UNIVERSITY OF ZAGREB
FACULTY OF ORGANIZATION AND INFORMATICS
VARAŽDIN

CECIIS

30th Central European Conference on Information and Intelligent Systems

international doctoral seminar IDS

Proceedings of the 11th International Doctoral Seminar

Organizers:
University of Zagreb
Faculty of Organization and Informatics

Slovak University of Technology in Bratislava
Faculty of Material Science and Technology in Trnava

STU MTF
SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA
FACULTY OF MATERIALS SCIENCE AND TECHNOLOGY IN TRNAVA

3rd October 2019
Varaždin, Croatia

Proceedings of the

11th International Doctoral Seminar (IDS 2019)

Varaždin, 3rd October 2019

ISBN 978-953-6071-67-8

Title: **Proceedings of the 11th International Doctoral Seminar (IDS 2019)**

Publishers: **Faculty of Organization and Informatics**
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
Faculty of Materials Science and Technology in Trnava
Slovak University of Technology in Bratislava
Jana Bottu 25, 917 01 Trnava, Slovakia

For the publishers: Neven Vrček and Oliver Moravčik

Editors: Diana Šimić
Maximilian Stremy
Nikolina Žajdela Hrustek

Technical editor: Igor Pihir

Available online: https://ceciis.foi.hr/doctoral-student-section/proceedings_IDS_2019.pdf

The CIP record is available in the computer catalog of the National and University Library in Zagreb under number 001042643.

ISBN: 978-953-6071-67-8

Welcome to the Proceedings of the IDS 2019, 11th International Doctoral Seminar. The Seminar offers doctoral students in the fields of Computer/Information and Material Sciences an opportunity to present their ideas, research proposals, preliminary results, problems and challenges and receive feedback and guidance from a panel of experts in their fields. It is co-organized by the Faculty of Organization and Informatics, University of Zagreb and Faculty of Materials Science and Technology in Trnava, Slovak University of Technology in Bratislava, as a special Doctoral Student Section within the 30th Central European Conference on Information and Intelligent Systems CECIIS 2019.

You will notice that contributions in these Proceedings are not uniformly formatted. In order to accommodate for different stages of doctoral study and research, we have three types of contributions – full research papers, short doctoral research papers, and research posters. Full research papers were presented both in an appropriate CECIIS section, and at the IDS. They are published in the CECIIS proceedings, and we reproduce here only their abstracts. Short doctoral research papers and posters were reviewed by the members of the international Program Committee, and their full contents are included in these Proceedings. The order of contributions in the Proceedings follows the order of presentations in the IDS 2019 Program to facilitate tracking of presentations.

We are grateful to all those who have supported the Seminar, and especially to our young authors. We hope that IDS will provide an encouraging and supportive setting for learning, exchanging ideas, and developing research and professional networks.

Diana Šimić, Chair Program Committee
Nikolina Žajdela Hrustek, Chair Organizing Committee

Sadržaj

Wonderware Production Line Model and MES System Creation for Data Analyzation

Andrea Václavová, Martin Batroň
Faculty of Materials Science and Technology in Trnava
Slovak University of Technology in Bratislava
Jana Bottu 25, 917 01 Trnava, Slovakia
{andrea.vaclavova, martin.barton}@stuba.sk

Abstract

The aim of this poster is to present the current status of research and to declare the further steps in working with Wonderware software, various products which it provides, and a fully automated production line. We have a production line model that consists of five stations. The production line is used for filling bottles with continuous (with the ability to mix accurate ratios) or discrete (with the ability to sort raw granulate) material according to the specified requirements. Next step in the production process is either storage or direct distribution of already filled bottles and possible re-emptying and prepare bottles for further reuse. To avoid losing data every time we shut down and restart production line, we had to create a communication and start saving the data. Currently, the signals from the PLC of the production line are linked and a communication between Wonderware the System Management Connsole (SMC) and ArchestrA Application production line is established. With this communications link, we have been able to create a complete production line model that includes all of the individual zones and their objects and its attributes in the ArchestrA IDE, and we are able to use this model to control the physical line while tracking the current values of each attribute. After deploying this model, we are storing all of this data with the Wonderware Historian Client, which offers various options for working with big data such as creating trends. Furthermore, it will be necessary to create a virtual production line model using the Wonderware HMI tool InTouch that will allow us to directly control and observe the individual signals of production line. Using all three Wonderware components (ArchestrA IDE, InTouch and Historian) we want to create a MES system, which can store data to database and later on we use this system to analyze data from production line.

Keywords: wonderware, integration, big data, data storage

Acknowledgments

This publication was created thanks to the support of the Ministry of Education, Science, Research and Sport of the Slovak Republic in the framework of the call for subsidy for the development project No. 002STU-2-1/2018 with the title „ STU as the Leader of the Digital Coalition

References

[1] Stenerson J, Deeg D: **Programing Siemens Step 7 (TIA Portal) Programing, a Practical Approach**, CreateSpace Independent Publishing Platform, 2015

[2] Tempest, **Production line description**, User manual, Tempest, 2014

[3] Wonderware, **Application Server**, Training manual, Achneider Electric Software, 2015

[4] Wonderware, **Wonderware Historian Industrial Data Management**, available at https://www.wonderware.com/industrial-information-management/historian/, Accessed 29th April 2019

[5] Wonderware, **Wonderware InTouch**, available at https://www.wonderware.com/hmi-scada/intouch/, Accessed 2nd May 2019

[6] Wonderware, **Wonderware Software**, available at https://www.wonderware.com, Accessed 29th April 2019

Optimization of Laser Welding Parameters of Copper to Austenitic Stainless Steel

Ema Tomčíková
Faculty of Materials Science and Technology in Trnava
Slovak University of Technology in Bratislava
J. Bottu 25, Trnava, 917 01 Slovakia
ematomcikova@gmail.com

Abstract

This paper deals with laser beam welding of dissimilar metals. Oxygen free copper C 10 600 and austenitic stainless steel AISI 304, with the thickness of 1 mm were proposed as experimental materials. Usually welding of dissimilar materials results in the formation of hard and brittle intermetallic compounds. Contrary, no intermetallic compounds are formed between iron and copper. Weld joint consisting of mentioned materials combination, offers high electrical and thermal conductivities and superior corrosion resistance. Thus, dissimilar metals joints are utilized in various fields of industry such as power generation, nuclear and heat exchanger. Generally welding of dissimilar materials is great challenge because of large differences in thermal properties. Usually the main problem is liquation cracking in heat affected zone of austenitic stainless steel, deteriorating mechanical properties of weld joint. It is necessary also to take into the consideration the high reflectivity of laser radiation by the surface of copper and its alloys. From this point of view utilization of solid-state disk laser for joining could be advantageous due to shorter wavelength. It is supposed that weld joint will have heterogeneous character from microstructural view. [1] Design of experiment was used in order to evaluate influence of welding parameters on geometry and mechanical properties of weld joints and optimization of welding parameters as well. Design of experiment will be performed in Minitab statistical software. Three factors namely laser power, welding speed and laser beam offset were selected. Each factor exhibited three different levels. The beam offset had important influence on mechanical properties of weld joints. Variations of mechanical properties will be associated with microstructure changes in weld metal. It is assumed that differences in laser beam offset will result in various intermixing of both metals. Furthermore, light microscopy, scanning electron microscopy and microhardness measurements will be used to analyse weld joint properties.

Keywords: laser beam welding, copper, stainless steel, optimization, design of experiment

Acknowledgments

This research is part of final paper at the Faculty of Materials science and technology in Trnava "Optimization of laser welding of copper to austenitic stainless steel", whose mentor is Ing. Miroslav Sahul, PhD..

References

[7] Antoine Mannucci, Iryna Tomashchuk, Vincent Vignal, Pierre Sallamand, Melanie Duband: **Procedia CIRP,** Université de Bourgogne Franche Comté, France, 2018.

Diagnostics of the Machines and Devices Based on the SIMATIC ProDiag

Michal Kopček, Miroslava Fačková
Faculty of Materials Science and Technology in Trnava
Slovak University of Technology in Bratislava
J. Bottu 25, 917 24 Trnava, Slovakia
{michal.kopcek, xfackova}@stuba.sk

Michal Kopček
Skartek s.r.o.
Vlčkovce 474, 919 23 Vlčkovce, Slovakia
m.kopcek@skartek.com

Abstract

The poster is focused on the diagnostics of machines and devices based on the SIMATIC ProDiag library for the machines produced by company Skartek, s. r. o. The ProDiag software library is used to diagnose machines and devices by configuring machine status tracking. The first part is dedicated to theoretical problem description and analysis. In order to be able to assess appropriateness of the ProDiag library a model solution was designed and implemented in conformance to the internal company programming standard. (Kopček, 2019). The next part of the work describes a sample example with configuration of supervisions, which serves as a demonstration for the implementation of the ProDiag library. The model is conceived as a simple assembly line with three workstations (WS) with palette transfer conveyor for palette transport between operators. The palette could be moved in the two directions by the conveyor. The palette could be also moved towards the operator and back on WS 1 and WS 3 using pneumatic actuators. Therefore, optical barriers are presented on these posts. The WS 1 serves for loading of the part to the jig on the palette by operator N°1. The WS 2 simulates the manual assembly process, which is irrelevant from the diagnostics point of view, therefore was not simulated. The WS 3 serves for unloading of the part from the line. The functions of the SIMATIC ProDiag software library are configured in a separate function block through both local and global supervisions. The ProDiag Library offers a configuration of five types of supervisors that are implemented and tested on a model. Operand, Interlock, Action, Reaction and Position supervision are used to supervise the normal execution of the work cycle. After implementation and testing phase of the model solution an evaluation of the utility of the ProDiag software library within the production of the Skartek company could be done. The comparison of alarm handling by conventional ways and by using the ProDiag library is presented at the end of this contribution.

Keywords: PLC, SIMATIC ProDiag, alarm, supervisions, maintenance

Acknowledgments

This publication was created thanks to the support of the Ministry of Education, Science, Research and Sport of the Slovak Republic in the framework of the call for subsidy for the development project No. 002STU-2-1/2018 with the title „ STU as the Leader of the Digital Coalition

References

[1] Fačková, M., (2019). **SIMATIC ProDiag based machine and plant diagnostics**, *in Slovak (Diagnostika strojov a zariadení na báze SIMATIC ProDiag)*, Diploma thesis, Supervisor: Kopček, M., Trnava: UIAM MTF STU. MTF-30179-73437

[2] Kopček, M., (2014). **Management of FAT for the process level control systems**, *in Slovak (Manažment FAT pre systémy riadenia procesnej úrovne)*, Habilitation thesis, Trnava: UIAM MTF STU.

[3] Kopček, M., et al. (2014). **Systematic approach to factory acceptance test planning,** In ESCAPE-24: 24th European symposium on computer aided process engineering. Part A.

Hungary, Budapest, 15-18 June 2014. 1. ed. Amsterdam: Elsevier, p. 1597-1602. ISBN 978-0-444-63456-6

[4] Kopček, M., (2019). **Standard programming guide v1.1.,** *in Slovak*, *(Štandard programovania v1.1., internal documentation)*, Trnava: Skartek

[5] NED, (2016). **Software PRODIAG***,* © Nuclear Engineering Division 2016 [cit. 2019-06-16], available at https://www.ne.anl.gov/codes/prodiag/

[6] Rothenberg, D.H., (2009). **Alarm Management for Process Control: A Best-Practice Guide for Design, Implementation, and Use of Industrial Alarm Systems,** Momentum press, ISBN-13: 978-1-60650-003-3

[7] Siemens, (2018). **Machine and Plant Diagnostics with ProDiag V4.1,** © Siemens AG 2019 [cit. 2019-06-16], available at https://support.industry.siemens.com/cs/ww/en/viev/109740151

[8] Strémy, M., et al. (2011). **Introduction to programmable logic controllers***, in Slovak, (Úvod do programovateľných logických automatov)*, Trnava: Qintec. ISBN 978-80-969846-9-5

Implementation of Digital Technologies in Smart Factory Processes

Larisa Hrustek, Martina Tomičić Furjan
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
{larisa.hrustek, martina.tomicic}@foi.hr

Abstract

Digital transformation implies changes in organizations through application of digital technologies to enhance existing or develop new processes. One of key features of digital technologies is the development of smart systems, whereby, when considering the manufacturing sector, traditional factories are transformed into smart factories. This article defines smart factories, explains main approaches to transform from traditional to smart and investigates the implementation of digital technologies in smart factory processes involved. The research aims to integrate and extend different approaches to transformation as well as to show how factories should use the digital enablers to benefit and improve their work.

Keywords: Smart factory, digital transformation, Industry 4.0

References

[1] Burke, R., Musoomeli, A., Laaper, S., Hartigan, M., Sniderman, B: **The smart factory – Responsive, adaptive, connected, manufacturing**. available at https://www2.deloitte.com/insights/us/en/focus/industry-4-0/smart-factory-connected-manufacturing.html, Accessed: 10th July 2019.

[2] Choi, JH., Jung, H., Lee, KW., Cho, ES. **Grand challenge: Runtime anomaly detection method in smart factories using machine learning on RDF event streams**. Proceedings of the 11th ACM International Conference on Distributed Event-Based Systems, 2017, pp. 304-309.

[3] Guo, Z., Zhang, K., Xin, H., Bi, M., He, H., Hu, W. **An Optical Access Network Framework for Smart Factory in the Industry 4.0 Era Supporting Massive Machine Connections.** 16th International Conference on Optical Communications and Networks (ICOCN), 2017.

[4] Ha, J., Kim, J., Park, H., Lee, J., Jo, H., Kim, H., & Jang, J. **A web-based service deployment method to edge devices in smart factory exploiting Docker**. International Conference on Information and Communication Technology Convergence: ICT Convergence Technologies Leading the Fourth Industrial Revolution, 2017, pp. 708-710.

[5] He, J., Jia, G., Han, G., Wang, H., Yang, X. **Locality-Aware Replacement Algorithm in Flash Memory to Optimize Cloud Computing for Smart Factory of Industry 4.0**. IEEE Access, Vol. 5, 2017, pp. 16252-16262.

[6] Huang, B., Wang, W., Ren, S., Zhong, RY., Jiang, J. **A proactive task dispatching method based on future bottleneck prediction for the smart factory**. International Journal of Computer Integrated Manufacturing, Vol. 32(3), 2019, pp. 1-16.

[7] Huang, D-C., Chen, C-Y., Lin, C-F., & Sze, J-R. **The Internet technology for defect detection system with deep learning method in smart factory**. 4th International Conference on Information Management, ICIM 2018, 2018, pp. 98-102.

[8] Ivanov, D., Dolgui, A., Sokolov, B., Werner, F., Ivanova, M. **A dynamic model and an algorithm for short-term supply chain scheduling in the smart factory industry 4.0**. Internationl Journal of production research, Vol. 54(2), 2016, pp. 386-402.

[9] Kang, S., Jeon, J., Kim, H.-S., & Chun, I. **CPS-based fault-tolerance method for smart**

factories. Automatisierungstechnik 2016, 64(9), 2016, pp. 750-757.

[10] Lederer, M., Betz, S., & Schmidt, W. **Digital transformation, smart factories, and virtual design - contributions of subject orientation**. ACM International Conference Proceeding Series, 2018, pp. 1-10.

[11] Lima-Monteiro, P., Parreira-Rocha, M., Rocha, AD., Barata Oliveira, J. **Big Data Analysis to Ease Interconnectivity in Industry 4.0-A Smart Factory Perspective**. Service Orientation in Holonic and Multi-Agent Manufacturing: Proceedings of SOHOMA 2016, 2017, pp. 237-245.

[12] Longo, F., Nicoletti, L., & Padovano, A**. Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context**. Computer & Industrian Engineering, Vol. 113, 2017, pp. 144 -159.

[13] Loucks, J., Macaulay, J., Noronha, A., & Wade, M. **Digital Vortex – How todays market leaders can beat disruptive competitors at their own game**. International Institute for Management Development, Switzerland, 2016

[14] Paelke, V. **Augmented Reality in the Smart Factory Supporting Workers in an Industry 4.0. Environment.** 19th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA, 2016.

[15] Park, J. **A Smart Factory Operation Method for a Smart Grid**. International Conference on Computers and Industrial Engineering: Soft Computing Techniques for Advanced Manufacturing and Service Systems, 2010.

[16] Park, D-J., Lee, J-H., Park, J-A., & Bae, H-S. **The method of structuring production logistics systems in smart factories of SMEs.** International Journal of Advanced Science and Technology, Vol. 118, 2018, pp. 119-130.

[17] Park, S. and Lee, SJ. **A Study on Worker's Positional Management and Security Reinforcement Scheme in Smart Factory Using Industry 4.0-Based Bluetooth Beacons**. Advances in Computer Science and Ubiquitous Computing, Lecture Notes in Electrical Engineering 421, 2017, pp. 1059-1066.

[18] Pei, F., Tong, Y., He, F., Li, D. **Research on design of the smart factory for forging enterprise in the industry 4.0 environment**. Mechanika, Vol. 23(1), 2017, pp. 146-152.

[19] Ren, G., Hua, Q., Deng, P., Yang, C., & Zhang, J. **A Multi-Perspective Method for Analysis of Cooperative Behaviors among Industrial Devices of Smart Factory**. IEEE Access, 2017, pp. 1-10.

[20] Suginouchi, S., Kokuryo, D., & Kaihara, T. (2017). **Value Co-creative Manufacturing System for Mass Customization: Concept of Smart Factory and Operation Method Using Autonomous Negotiation Mechanism**. The 50th CIRP Conference on Manufacturing Systems 63, 2017, pp. 727-732.

[21] Schwab, K. **The fourth industrial revolution**. Portfolio penguin, United Kingdom, 2017.

[22] Trinks, S., and Felden, C. **Edge Computing architecture to support Real Time Analytic applications A State-of-the-art within the application area of Smart Factory and Industry 4.0**. IEEE International Conference on Big Data (Big Data), 2018, pp. 2930-2939.

[23] Wan, J., Tang, S., Li, D., Imran, M., Zhang, C., Liu, C., Pang, Z. **Reconfigurable Smart Factory for Drug Packing in Healthcare Industry 4.0**. Transactions on Industrial Informatics, 2019, pp. 1-9.

[24] Wan, J., Xia, M., Hong, J., Pang, Z., Jayaraman, B., & Shen, F. **IEEE ACCESS Special section editorial: Key technologies for smart factory of Industry 4.0**. IEEE-INST electrical electronics engineers, Vol. 7, 2019, pp. 17969-17974.

[25] Wang, S., Wan, J., Zhang, D., Li, D., & Zhang, C. **Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination**. Computer Networks 101, 2016, pp. 158–168.

[26] Zhang, Z., Li, X., Wang X., & Cheng H. **Decentralized Cyber-Physical Systems: A Paradigm for Cloud-Based Smart Factory of Industry 4.0**. Springer International Publishing, 2017, pp. 127-171.

Challenges, Issues, Barriers and Problems in Digital Transformation – Systematic Literature Review

Ana Kutnjak, Igor Pihir
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
{ana.kutnjak, igor.pihir}@foi.hr

Abstract

Digital Transformation (DT), an inevitable and necessary part of today's organizations lifecycle, is a relevant topic in business circles. Well strategically designed, it provides opportunities for organizations to achieve momentum in their business, thereby increasing organizational efficiency. Digital technologies enable digital transformation with full involvement of all organizational resources, working together to transform processes, business models and achieve new value for end-users and organization.

Because of DT process complexity, this research presents systematic literature review to systemise challenges, issues, barriers and problems that influence digital transformation and presents a challenge for every company that carry out the DT process.

Keywords: Digital transformation, challenges, issues, barriers, problems, systematic literature review.

References

[1] Achatz, R: **Digital Transformation at thyssenkrupp: Challenges, Strategies and Examples**, International Conference on Advanced Information Systems Engineering, Springer, Cham, 2017, pp. 3-12.

[2] Afonasova, M.A: **Digital Transformation of the Entrepreneurship: Challenges and Prospects**, available at: https://www.abacademies.org/articles/digital-transformation-of-the-entrepreneurship-challenges-and-prospects-7646.html, Accessed: 28th July 2019.

[3] Al-Ruithe, M, Benkhelifa, E, Hameed, K: **Key Issues for Embracing the Cloud Computing to Adopt a DT: A study of Saudi Public Sector**, available at https://www.sciencedirect.com/science/article/pii/S1877050918305076, Accessed: 28th July 2019.

[4] Boronos, V, Plikus, I, Aleksandrov, V, Antoniuk, N: **Digital transformation of Ukraine: challenges of theory and practice in implementation of digital quality of life**, Economic Annals-XXI, 2018, pp. 38-43.

[5] Cambridge Dictionary: **Barrier**, available at https://dictionary.cambridge.org/dictionary/english/barrier, Accessed: 28th July 2019.

[6] Cambridge Dictionary: **Challenge**, available at https://dictionary.cambridge.org/dictionary/english/challenge, Accessed: 28th July 2019.

[7] Catarino, J, Rosa, I, Mira da Silva, M: **Defining the Chief Digital Officer Using COBIT 5**, available at http://www.isacajournal-digital.org/isacajournal/2018_volume_6/MobilePagedArticle.action?articleId=1438649#articleId1438649, Accessed: 28th July 2019.

[8] Heavin, C, Power, D: **Challenges for digital transformation – towards a conceptual decision support guide for managers,** Journal of Decision Systems, 2018, pp. 38-45.

[9] Henriette, E, Feki, M, Boughzala, I: **Digital transformation challenges**, Proceedings of Tenth Mediterranean Conference on Information System (MCIS), Association for Information Systems, 2016, pp. 1-7.

[10] Horlacher A, Hess, T: **What Does a Chief Digital Officer Do? Managerial Tasks and Roles of a New C-level Position in the Context of Digital Transformatio**, Proceedings of 49th Hawaii

International Conference on System Sciences, IEEE Computer Society, 2016, pp. 5126-5135.

[11] Hossain, S. A: **Blockchain Computing: Prospects and Challenges for Digital Transformation**, Proceedings of 6th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), AIIT, Amity University Uttar Pradesh, 2017, pp. 61-65.

[12] Hrustek, L, Tomičić Furjan, M, Pihir, I: **Influence of Digital Transformation Drivers on Business Model creation**, Proceedings of the 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics 2019, pp. 1509-1513.

[13] Ianenko, M, Ianenko, M, Huhlaev, D, Martynenko, O: **Digital transformation of trade: problems and prospects of marketing activities**, Proceedings of Materials Science and Engineering Conference, IOP Publishing, 2018, pp. 1-5.

[14] Karim, H. S. B. H. A: **Digital transformation of libraries in Brunei Darussalam: addressing the sustainability issues of VILIS Brunei**, Emerald Insight, 2004, pp. 184-193.

[15] Khitskov, E, Veretekhina, S, Medvedeva, A, Mnatsakanyan, O, Shmakova, E, Kotenev, A: **Digital Transformation of Society: Problems Entering in the Digital Economy**, Eurasian Journal of Analytical Chemistry, 2017, pp. 855-873.

[16] Knop, J: **Chemistry 4.0 Challenges and Solutions for the Digital Transformation**, Croatica Chemica Acta, 2016, pp. 397-402.

[17] Kutnjak, A, Pihir, I, Tomičić Furjan, M: **Digital Transformation Case Studies Across Industries – Literature Review**, Proceedings of the 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics 2019, pp. 1498-1503.

[18] Kutnjak, A, Križanić, S, Pihir, I: **Educational and practical view of knowledge, skills and experience needed by a Chief Digital Officer**, Proceedings of the 11th International Conference on Education and New Learning Technologies (EDULEARN 2019), pp. 5711-5718.

[19] Limani, Y, Stapleton, L, Groumpos, P.P: **The Challenges of Digital Transformation in Post-Conflict Transition Regions: Digital Technology Adoption in Kosovo**, Proceedings of 18th IFAC Conference on Technology, Culture and International Stability TECIS, 2018, pp. 186-191.

[20] Maltese, V: **Digital Transformation Challenges for Universities: Ensuring Information Consistency Across Digital Services**, Cataloging & Classification Quarterly, 2018, pp. 592-606.

[21] Mazzone, D: **Digital or Death: Digital Transformation – The Only Choice for Business to Survive, Smash, and Conquer**, Smashbox Consulting Inc, 2014.

[22] Mendhurwar, S, Mishra, R: **Integration of social and IoT technologies: architectural framework for digital transformation and cyber security challenges**, Enterprise Information Systems, 2019, pp. 1-20.

[23] Panichkina, M, Sinyavskaya, I, Shestova, E: **Challenges of Professional Adaptation of University Graduates in Response to the Economics' Digital Transformation**, Proceedings of the XVII Russian Scientific and Practical Conference on Planning and Teaching Engineering Staff for the Industrial and Economic Complex of the Region (PTES), 2018, pp. 44-46.

[24] Pereira, J.L, Belo, O, Ravesteijn, P: **Special Issue on Big Data and Digital Transformation**, Journal of Grid Computing, 2018, pp. 531-533.

[25] PhraseMix: **What's the difference between "problem", "trouble", and "issue"?** available at https://www.phrasemix.com/answers/whats-the-difference-between-problem-trouble-and-issue, Accessed: 28th July 2019.

[26] Rodrigues, L.S: **Challenges of Digital Transformation in Higher Education Institutions: A brief discussion**, 32th IBIMA Conference, 2017, pp. 1-4.

[27] Salem, F: **Building a Smart City: Overcoming the Challenges of DT – The Case of „Smart Dubai“**, available at: https://books.google.hr/books?id=J7IpDwAAQBAJ&pg=PA139&lpg=PA139&dq=10.1108/S2048-75762017000006006&source=bl&ots=PqmlZBPvCx&sig=ACfU3U14YkdApD3U8qsxIlwb2KT1ahDI7w&hl=hr&sa=X&ved=2ahUKEwjMrvvMt6fjAhVN3KQKHTYzCgIQ6AEwAHoECAAQAQ#v=onepage&q=10.1108%2FS2048-75762017000006006&f=false, Accessed: 28th July 2019.

[28] Schwanholz, J, Graham, T: **Digital Transformation: New Opportunities and Challenges for Democracy?** Managing Democracy in the Digital Age: Internet Regulation, Social Media Use and Online Civic Engagement, 2018, pp. 1-7.

[29] Shafiee Nahrkhalaji, S, Shafiee, S, Shafiee, M, Hvam, L: **Challenges of Digital Transformation: The Case of the Non-profit Sector**, Proceedings of 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2018, pp. 1245-1249.

[30] Stark, R, Damerau, T, Lindow, K: **Industrie 4.0—Digital Redesign of Product Creation and Production in Berlin as an Industrial Location**, The Internet of Things, 2018, pp. 171-186.

[31] Tiersky, H: **Navigating digital transformation,** available at https://www.cio.com/article/3179607/5-top-challenges-to-digital-transformation-in-the-enterprise.html, Accessed: 28th July 2019.

[32] Tumbas, S, Berente, N, von Brocke, J: **Three Types of Chief Digital Officers and the Reasons Organizations Adopt the Role**, MIS Quarterly Executive, 2017, pp. 121-134.

[33] Vasiljeva, T, Lukanova, K: **Commercial banks and Fintech companies in the digital transformation: challenges for the future**, Journal of Business Management, 2016, pp. 25-33.

[34] Yli-Huumo, J, Smolander, K: **Changes and challenges of technical debt and its management during ongoing digital transformation**, Ninth International Workshop on Managing Technical Debt, 2017, pp. 1-4.

[35] Zahariaa, S.E, Pietreanub, C.V: **Challenges in airport digital transformation**, Transportation Research Procedia, 2018, pp. 90-99.

A Public Policy Readiness for e-Inclusion of 54+

Robertina Zdjelar
PhD Student on Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
rzdjelar@foi.hr

1. Introduction

The area of research is the readiness of public policy for e-inclusion of 54+. Generally, the rate of e-inclusion of 54+ is low, especially for the population 60+ globally and in the European Union as it is documented by EUROSTAT [6]. The motivation for this research lays on low rate of e-Inclusion of 54+ in the last decade and on many scientific and professional articles and books that relates with adult's behaviour, quality of life, e-inclusion, psychology, social psychology and public policies. The Croatian National statistics about digital literacy research conducted by Census 2011 presents also the low level of digital literacy rate of 54+ in Croatia [9]. Review of previous relevant research of e-inclusion of elderly have been presented in articles [15], [16], [17]. The quality of life is wide term in general [11], [12], [13]. Considering the elderly everyday life, many problems caused by physical barriers can be avoided if individuals get and use appropriate digital literacy skills [11], [12], [13], [17]. According to the [14] "E-inclusion or digital divide: an integrated model of digital inequality" public policy [7] is one of the forces of the macro causes that determines the resources, as meso causes to assure access as a micro cause. *Considering presented arguments the research problem* is how the public policies have influence to digital literacy and e-inclusion of 54+ rate. Methodology of creating public national policies and regional policies in Croatia are regulated by law [8] and it is justified with EU law. This research is oriented to readiness of public policies for e-inclusion of 54+ based on referenced models: 1) Proposed General Theoretical Conceptual Model of e-Inclusion [18] and 2) An Integrated model of digital inequality [14].

2. Research Methods

The shortcomings from previous research e-Inclusion of 54+ and facts that present the actual rate of e-Inclusion of 54+, public policy documents and attitudes of policy makers will be used as a source for core data to set a new theory by using Grouded Theory method.

Activities	Research questions	Research objectives	Methods
Define the Research Problem	What are the results (rate) of e-Inclusion of 54+? What kind of causes make the public policy of e-Inclusion of 54+ rate?	To synthesize the results of previous researches on the topic e-Inclusion of 54+ and public policies. Set the terminology; determine referenced e-Inclusion models	Literature Review; Extracting data from EUROSTAT database; Extracting data from CENSUS RH 2011 database
Define the categories and subcategories based on analyses of artefacts (public policy documents)	How to measure public policy readiness for e-Inclusion of 54+ in accordance with referenced models?	To find the categories and subcategories that participated in the model that represent the readiness of public policies for e-Inclusion of 54+	Grounded Theory [1], [2] applied on: - public policy strategic documents; - data collected from Focus Group and Expert Judgment (interview)

Define the relationships between categories and subcategories of the model	How to determine the relationship between the categories and subcategories based on previous step of the research	To determine the readiness assessment model for public policies e-Inclusion of 54+ and to create recommendations for improving strategic planning of e-inclusion of 54+ for policy makers	Grounded Theory [1], [2] applied on: - public policy strategic documents; - data collected from Focus Group and Expert Judgment (interview)
Evaluation of the results	How to evaluate the results of readiness assessment model for public policies e-Inclusion of 54+?	To evaluate the results of a created readiness assessment model for public policies e-Inclusion of 54+	Expert evaluation with Card sorting method

3. Results and discussion

Completed researches that author have done are: **(1)** *r*esearch of e-inclusion of 54+ rate and inactive individuals' rate in EU countries (EUROSTAT database of e-society [6]); the data were extracted in additional research material that contains many comparations of data by many criteria in the period of 2007 – 2018. ; **- (2)** the e-inclusion of 54+ and digital literacy have been very often researched as skills and opportunities, also from the aspect of user readiness and attitudes about using Information and Communication technologies, Internet and e-services. Many projects have been performed in practice; the awareness of low rate of e-Inclusion of 54+ has been communicated to publicity. Author has communicated the results of mentioned analyses on conferences and the papers have been published in proceedings books referenced in WoS and on Research gate [15] [16] [17]. ; **(3)** – as the referenced e-inclusion model, for this research, has been accepted Proposed General Theoretical Conceptual Model of e-Inclusion [18] and an integrated model of digital inequality [14].
The ongoing research will consist of detecting public policy strategic documents on UN and EU level, as well as national level of EU countries and regional level of Croatian regional counties (digital transformation; sustainable development policy; active ageing policy). The Grounded theory method will be applied on mentioned documents and notes from interviews. *The second phase* of research will consists of organizing the Focus groups of policy makers as group interview. That source will provide the data that will be used to determine the impact of policy makers' attitudes about e-Inclusion of 54+ on the created policy documents. The expert for creating public policy will be engaged to make an expert judgement of the proposed categories and subcategories as well as for evaluation of the composition that will be made at the end. The scientific contribution of the research will be a proposed assessment model of public policy readiness for e-inclusion of 54+.
The expected social contribution of research will be the recommendations for improving strategic planning of e-inclusion of 54+ for policy makers.

4. Conclusion

The *importance* of the research is to detect the readiness of public policy for e-Inclusion of 54+ and to formulate the recommendations for the actions that need to be taken in further strategic planning measures for raising the rate of e-Inclusion of 54+. The *Value of proposed research* is in creating the new active environmental opportunities to input e-inclusion of 54+ in strategy document for further 6 years period by using suggested recommendations.

5. References

[1] Charmaz, K., Constructing Grounded Theory A Practical Guide through Qualitative Analysis, SAGE, 2006.

[2] Creswell, J.W. Research design Qualitative, Quantitative and Mixed Methods Approaches, London, 2009.

[3] Dzhusupova, Z., Shareef, M., Ojo, A., Janowski, T. Methodology for e-Government Readiness Assessment – Model, Instruments, Implementation, Centre for Electronic Governance, United Nation University, International Conference on Society and Information Technologies (ICSIT 2010), Florida, USA, January 2010

[4] European Commission, Digital Agenda 2020, The Digital Skills and Jobs Coalition, (https://ec.europa.eu/digital-single-market/en/digital-skills-jobs-coalition, accessed December 1st 2016)

[5] European Commission, MIREIA - Measuring the Impact of eInclusion Actors, Measuring the Impact of eInclusion Actors, Exploratory Study on Methods used to measure the ICT-mediatied Social Impact of Grassroots Organizations, JCR Technical Report EUR 25494EN, EU 2012. European Commission, The

Digital Economy and Soceiety Index (DESI) 2019. How digital is your country? Europe improves but still needs to close digital gap https://ec.europa.eu/digital-single-market/desi, accessed July 25th 2019.)

[6] EUROSTAT, Digital economy and society, http://ec.europa.eu/eurostat/data/database, accessed February 4th 2018.

[7] Mukosa, F., Mweemba, B., Katebe, M. What is Public Policy? How is it related to the law?, Imperial Journal of Interdisciplinary Research (IJIR), Vol-3, Issue-9, 2017 (IJIR), ISSN: 2454-1362, http://www.onlinejournal.in

[8] Narodne novine, Zakon o sustavu strateškog planiranja i upravljanja razvojem Republike Hrvatske, broj 123/17

[9] Popis stanovništva 2011. u Republici Hrvatskoj,(http://www.dzs.hr/, accessed January 10th 2017)

[10] Tierney, W.G., Clemens, R.F., Qualitative Research and Public Policy: The Challenges of Relevance and Trustworthiness, Center for Higher Education Policy Analysis, University of Southern California, 3470 Trousdale Parkway, WPH 701D, Los Angeles, CA 90089

[11] United Nations. (2015). *Transforming our World: The 2030 Agenda for Sustainable Development.* accessed August 1st 2018 from https://sustainabledevelopment.un.org/post2015/transformingourworld

[12] United Nations. (2017). *Population facts: Population ageing and sustainable development.* Department of Economic and Social Affairs, Population Division, No. 2017/1, accessed 20.01.2019 from http://www.un.org/en/development/desa/population/publications/pdf/popfacts/PopFacts_2017-1.pdf

[13] Vincent, J.A. *Older People and Sustainable Development*. Department of Sociology, University of Exeter, UK, accessed January 15th 2019 from https://www.eolss.net/Sample-Chapters/C13/E1-20-04-01.pdf

[14] Yu, B., Ndumu, A., Mon, L.M., Fan, Z. (2018) "E-inclusion or digital divide: an integrated model of digital inequality", Journal of Documentation, Vol. 74 Issue: 3, pp.552-574

[15] Zdjelar R. (2019.) "WHO CARES ABOUT DIGITAL LITERACY OF 54+ CITIZENS?!", CEE e| DEM and e|GOV Days 2019, Budapest, May, 2019.

[16] Zdjelar, R., Kelemen, R. (2018). *The Smart Cities are implemented – Are Citizens "Smart" Also.* 6th Smart Cities Conference, 2018 Bucharest, Romania

[17] Zdjelar, R., Žajdela Hrustek, N., Sumpor, M. (2019) Sustaintable development and active ageing in EU countries – bridges and gaps, ESD-Conference, Rabat, 2019

[18] Žajdela Hrustek, N. Multidimensional and multiperspective approach for monitoring e-inclusion, Doctoral thesis, Faculty of Organization and Informatics, Varaždin, 2015

Implementation of the Cybersecurity for Medical Devices through the Healthcare Enterprise Reference Architecture (HERA) Framework

Nadica Hrgarek Lechner
Faculty of Organization and Informatics
Pavlinska 2, 42 000 Varaždin, Croatia
nhrgarek@foi.hr

1. Introduction

The ability of information and communications technology to improve patient care and experience, and to provide patients and medical professionals with valuable health information is leading to its increasing use in the healthcare industry. Medical devices are systems consisting of many components (e.g., hardware, firmware, software) that build complex architectures. Such systems are becoming increasingly connected using wired or wireless network capabilities. Connected (implantable) medical devices that have a direct impact on patients' health and well-being pose a higher security risk and have increased need for cybersecurity. The next decade is likely to witness a considerable rise in connectivity of medical devices as well as increase in security breaches and cybersecurity threats. From a security point of view, another challenge in healthcare are older medical devices that are still in use. Such devices were very often designed without or with minimum security controls.

Design and development of medical devices involves a significant number of regulations and standards to demonstrate that they meet legal requirements in order to ensure they are safe, secure, and effective throughout the entire life cycle. Legal requirements include quality management system, risk management system, usability engineering, biocompatibility, sterility, medical testing, cybersecurity, privacy, etc. Hrgarek Lechner (2017) provides a brief overview of cybersecurity regulations and standards for medical devices. The increasing amount of guidance documents, national laws and regulations, and standards in different countries to strengthen cybersecurity requirements is a challenging area in the field of medical devices.

Poor cybersecurity implementation can affect patient health, unintentionally expose patient data, and lead to a significant adverse impact on the business. Medical device manufacturers operating in global context need a solution to implement a cybersecurity risk management program that fulfils cybersecurity requirements throughout the design and development process of medical devices that are subject to cybersecurity. Implementing such a program must address patient's safety and privacy, and product security. Typically, the applicable guidance documents (FDA, 2014; FDA, 2016; Health Canada, 2019; TGA, 2019) recommend the use and adoption of the voluntary NIST Cybersecurity Framework (NIST, 2018), diverse security standards and privacy regulations. In addition, the guidance documents provide examples of security risk controls that should be considered for medical devices. Currently there is no international cybersecurity standard that is solely focused on the medical device industry. For example, the payment card industry uses a set of security standards to protect cardholder payment data (PCI Security Standards Council, 2018). Another regulated industry like automotive is developing a new international standard for automotive cybersecurity engineering (ISO/SAE, 2019). In analogy to other regulated industries, International Electrotechnical Commission (IEC) is developing a new technical report IEC TR 60601-4-5 and a new standard IEC WD 80001-5-1 to support medical device manufacturers. However, adding more guidance documents and standards increases the complexity of the evolving medical device cybersecurity ecosystem.

Our ongoing research is guided by the following preliminary research questions:

1. Is there a need to develop a new approach/solution to address the identified research problem?
2. What methodologies and standards exist to solve the research problem?
3. What industrial solutions exist to solve the research problem in the medical device domain?
4. What solutions from other industries may be used to solve the research problem?
5. What is missing in existing methodology during design and development of medical devices?
6. Is the solution going to be developed from scratch or extend/enrich an existing solution?
7. If the solution is going to or extend/enrich an existing solution, what needs to be enriched?

The purpose of this paper is to explore and briefly describe ideas through which the Healthcare Enterprise Reference Architecture (HERA) framework could be used/adapted to address challenges and for successful

implementation of a cybersecurity risk management program that is an integral part of design and development processes of medical devices.
We expect that the research is going to lead to many small contributions. The research will contribute to the existing literature by creating new knowledge and may improve research and practice. Our research has a great potential to classify many regulations and standards to overcome complexity and to define a secure development life cycle for the medical device domain. In addition, by performing a literature review of previous research we will identify papers where other researchers see solutions to solve the research problem.

2. Research methods

After defining the research problem and preliminary research questions in section 1, we started by performing a literature review of previous research to identify papers that could be relevant to solve the research problem. While there are many papers about enterprise architecture frameworks listed in section 3, to the authors' knowledge, no publications are available in the literature that investigate the applicability of the HERA framework in the medical device domain for cybersecurity purposes. To address the preliminary research questions, we are going to develop a questionnaire with different categories of questions related to cybersecurity of medical devices (e.g., device type, network capability, risk type, etc.). The questionnaire will be used over a six months period to collect qualitative and quantitative data during security risk assessment meetings at one medical device manufacturer that has legacy devices and devices at different development stages which are subject to cybersecurity. Collected research data will be anonymized, analysed, and interpreted. The research will be limited to the premarket life cycle phase of a medical device and future research should concentrate on the postmarket phase.

3. Results and discussion

The purpose of our ongoing preliminary research is to discuss one appropriate solution to implement a robust cybersecurity risk management program for medical devices. Such a program must implement cybersecurity regulations and standards and integrate security risk management activities. We thus considered the following three solutions: 1) workflow-oriented, b) process-oriented, and c) architecture-oriented. As a first step we took a workflow-oriented solution into consideration. Such a solution can be based on a rule-based expert system and provide a service to medical device manufacturers. For example, when a new cybersecurity standard or guidance is published, the knowledge base of the expert system would be upgraded with new rules and presented to the user through the user interface to become aware of new cybersecurity requirements. The workflow-oriented solution is not suitable to solve identified research problem, because the problem is stratified and not linear. As a next step, we considered a process-oriented solution because the medical device manufacturers have quality management systems that follow the process approach. A process-oriented solution consists of activities and the instructions describing what should be done, responsibilities who should do something and information objects that are needed to perform an activity. Medical device cybersecurity is a process that has to be integrated into the product development life cycle as part of a quality management system (Hrgarek Lechner, 2018). The process is not isolated, has many interfaces to the other processes within the organization and beyond it, and shall be considered at a system level. Therefore, we decided to further explore the applicability of an architecture-oriented solution that is based on a reference model. A reference model is a standard decomposition of a known problem into parts that cooperatively solve the problem (Bass et al., 2003). Reference models have many advantages: they are generic and provide a framework to represent properties of elements in a domain of interest through several layers, they can be integrated into a corporate wide view, used in defining system requirements, training and change management, and as reference guides. There is a number of enterprise architecture frameworks: Zachman framework (Zachman, 1987), The Open Group Architecture Framework (TOGAF) (The Open Group, 2018b), Federal Enterprise Architecture Framework (FEAF) (FEAF, 2013), Treasury Enterprise Architecture Framework (TEAF), etc. Sajid & Ahsan (2014) explored the applicability of the Zachman, TOGAF, FEAF, and TEAF frameworks to healthcare organizations. The authors propose a framework for creating integrated healthcare information enterprise architecture model based on the TOGAF framework. Our hybrid approach is based on the HERA framework (The Open Group, 2018a) that is suitable for the development of a reference architecture for a healthcare company. A snapshot of the HERA framework was published in April 2018 (The Open Group, 2018a). We aim to explore the applicability of the HERA framework for implementation of a cybersecurity risk management program within the medical device domain. Based on the findings, we are going to propose an architecture-oriented solution through which a cybersecurity risk management program can be implemented by a medical device manufacturer using the HERA framework.

4. Conclusion

This is an original research work and the paper has not been submitted to other journal for review and possible publication. The originality of our approach is that we are going to address cybersecurity in a model through

processes (e.g., planning, requirements, design, development, production), activities (e.g., security risk assessment, threat modelling, security testing, etc.), tool support during the activities, and roles (e.g., security architect, security tester, external security consultant, etc.) during the premarket life cycle phase of medical devices that are subject to cybersecurity. We expect that the defined model could be used in practice by medical device manufacturers.

5. References

[1] Bass, L., Clements, P., & Kazman, R. *Software Architecture in Practice*. Addison-Wesley, Boston, 2003.

[2] FDA, Content of Premarket Submissions for Management of Cybersecurity in Medical Devices – Guidance for Industry and Food and Drug Administration Staff, Food and Drug Administration, 2014. (https://www.fda.gov/media/86174/download accessed 01.09.2019.)

[3] FDA, Postmarket Management of Cybersecurity in Medical Devices – Guidance for Industry and Food and Drug Administration Staff, Food and Drug Administration, 2016. (https://www.fda.gov/media/95862/download accessed 01.09.2019.)

[4] FEAF, Federal Enterprise Architecture Framework Version 2, 2013. (https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/egov_docs/fea_v2.pdf accessed 09.09.2019.)

[5] Health Canada, Guidance Document: Pre-market Requirements for Medical Device Cybersecurity, Ottawa, 2019. (https://www.canada.ca/content/dam/hc-sc/documents/services/drugs-health-products/medical-devices/application-information/guidance-documents/cybersecurity-guidance.pdf accessed 01.09.2019.)

[6] Hrgarek Lechner, N. An Overview of Cybersecurity Regulations and Standards for Medical Device Software. In *Proceedings of the Central European Conference on Information and Intelligent Systems*, pages 237-249, Varaždin, 2017.

[7] Hrgarek Lechner, N. Developing a Compliant Cybersecurity Process for Medical Devices. In *Proceedings of the Central European Conference on Information and Intelligent Systems*, pages 197-204, Varaždin, 2018.

[8] International Electrotechnical Commission, IEC TR 60601-4-5, Medical electrical equipment – Part 4-5 Guidance and interpretation – Safety related technical security specifications for medical devices (https://www.iec.ch/dyn/www/f?p=103:14:13410394850840::::FSP_ORG_ID, FSP_LANG_ID:22678,25 accessed 07.09.2019)

[9] International Electrotechnical Commission, IEC WD 80001-5-1, Application of risk management for IT-networks incorporating medical devices – Part 5-1: Activities in the product lifecycle (https://www.iso.org/standard/76097.html accessed 07.09.2019)

[10] ISO/SAE, ISO/SAE CD 21434 Road Vehicles – Cybersecurity Engineering, 2019. (https://www.iso.org/standard/70918.html accessed 01.09.2019.)

[11] NIST, Framework for Improving Critical Infrastructure Cybersecurity, National Institute of Standards and Technology, 2018. (https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf accessed 01.09.2019.)

[12] PCI Security Standards Council, Payment Card Industry (PCI) Data Security Standard, 2018. (https://www.pcisecuritystandards.org/documents/PCI_DSS_v3-2-1.pdf accessed 01.09.2019.)

[13] Sajid, M., & Ahsan, K. Enterprise Architecture for Healthcare Organizations. *World Applied Sciences Journal*, 30(10):1330-1333, 2014.

[14] TGA, Medical device cyber security guidance for industry, Therapeutic Goods Administration, 2019. (https://www.tga.gov.au/sites/default/files/medical-device-cyber-security-guidance-industry.pdf accessed 01.09.2019.)

[15] The Open Group, Healthcare Enterprise Reference Architecture (HERA), Berkshire, 2018a. (https://publications.opengroup.org/s182 accessed 01.09.2019.)

[16] The Open Group, The TOGAF® Standard, Version 9.2 (Evaluation Copy), 2018b. (https://publications.opengroup.org/i182 accessed 08.09.2019.)

[17] Zachman, J. A. A Framework for Information Systems Architecture. *IBM Systems Journal*, 26(3):276-292, 198

Environmental Sustainability in Digital Marketing of Automotive Industry

Mária Haladová, Natália Nováková, Miloš Čambál, Jana Šujanová
Institute of Industrial Engineering and Management,
Faculty of Material Science and Technology of STU
Ulica Jána Bottu č. 2781/25 917 24 Trnava, Slovakia
{maria.haladova, natalia.novakova, milos.cambal} @stuba.sk, jsujanova@gmail.com

Abstract

The main aim of this article is to analyze and describe the evolution of environmental sustainability in marketing strategy of two companies, namely Volkswagen and PSA Group for the previous three years.
The first part of the research defines sustainability with a focus on an environmental aspect. The second part forms research questions, research hypotheses and describes researched sample in particular publications published on the main websites of both companies. The third part of this work represents results of this research using visual graphs. The last part evaluates findings and importance of information influence in the context of environmental sustainability

Keywords: sustainability, environmental sustainability, sustainability marketing, sustainability trend

Acknowledgments
„This publication was created thanks to the support of the Ministry of Education, Science, Research and Sport of the Slovak Republic in the framework of the call for subsidy for the development project No. 002STU-2-1/2018 with the title „ STU as the Leader of the Digital Coalition“

References

[1] Belz F., Peattie K. (2012). **Sustainability Marketing: A Global Perspective,** 2nd Edition. Wiley (2012). p.31 ISBN: 978-1-119-96619-7.

[2] Charter M., Peattie K., Ottman J., Polonsky MJ. (2006). **Marketing and sustainability,** available at www.cfsd.org.uk/smart-know-net/links/smart-know-net.pdf, Accessed: 15th September 2019.:

[3] Goodland R., Daly H. (1996). Environmental Sustainability: Universal and non-negotiable. Ecological Applications, Vol. 6, No. 4 (Nov., 1996), pp. 1002-1017. Published by: Ecological Society of America

[4] Makortoff, K. (2015). Volkswagen cut from top sustainability index, available at https://www.cnbc.com/2015/09/29/volkswagen-cut-from-dow-jones-sustainability-ranking.html?fbclid=IwAR2y-fcEKlkzDJK-RKtTyZIHJH_aFuGC7Z9Jq_IxdybxqZHQ-VI7WowNtnA, Accessed: 15th September 2019.:

[5] Kumar V., Rahman Z., Kazmi A., Goyal P. (2012). Evolution of sustainability as marketing strategy: Beginning of new era. Proceedings of the International Conference on Emerging Economies - Prospects and Challenges (ICEE-2012). available at www.sciencedirect.com, Accessed: 15th September 2019:

[6] Media. (2019), available at https://www.volkswagenag.com/en/media.html, Accessed: 15th September 2019:

[7] Noo-urai N., Jaroenwisan K. (2016). Sustainability marketing: A Changing of Marketing Concept lead to Sustainable Business. International Journal of Business and Social Science. Vol. 7, No.4., available at https://ijbssnet.com/journals/Vol_7_No_4_April_2016/13.pdf, Accessed: 15th September 2019:

[8] OECD. (2002). Towards Sustainable Household Consumption Trends and Policies in OECD Countries. Paris: OECD., available at https://read.oecd-ilibrary.org/environment/towards-

sustainable-household-consumption_9789264175068-en#page3, Accessed: 15th September 2019.

[9] PSA Group. (2019). Press Releases, available at https://media.groupe-psa.com/en/psa-peugeot-citro%C3%ABn/press-releases/group/function.include?page=35, Accessed: 15th September 2019:

[10] PSA Group. (2019). Pacific Strategies Assessments, available at https://media.groupe-psa.com/en, Accessed: 15th September 2019.:

[11] Sustainable Development goals. (2015). About Sustainable development goals, available at https://www.un.org/sustainabledevelopment/sustainable-development-goals, Accessed: 15th September 2019.:

[12] PSA Group. (2016). The PSA Group is included in the NYSE Dow Jones Sustainability Index for the first time, available at https://media.groupe-psa.com/en/press-releases/group/psa-group-included-nyse-dow-jones-sustainability-index?fbclid=IwAR1vl5wdhxb1X7bU4kHHbQww70GsW9BUrlQwv3_1-a71NV7uilJiKDyVrgk, Accessed: 15th September 2019.:

[13] World Commission on Environment and Development (WCED) (1987). Our Common Future: Brundtland Report, available at http://mom.gov.af/Content/files/Bruntland_Report.pdf, Accessed: 15th September 2019.:

Framework for Quality Assessment of Open Datasets

Barbara Šlibar
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42 000 Varaždin, Croatia
bslibar@foi.hr

1. Introduction

The value which can be gained from open data represents an innovation potential for individuals as well as for private or public organizations. Although, open data enables a vast innovation potential for those subjects, there is uncertainty about their quality. Hence, the questionable quality is a threat of value that can be generated from such data.

Initial literature review reveals that there is a lot of researches which investigate data quality, but only few of them are focused on the open data. Even in those researches which are concerned with quality of open data and in which the quality frameworks for open data are proposed, are mostly directed to quality assessment of open data portals and not to quality assessment of datasets published on them [1]–[3]. Equal quality dimensions are variously called in different studies which directly affects the understanding and solving quality problems. Thus, Neumaier, Umbrich and Polleres (2016) included dimension existence into their quality framework, while Vetrò et al. (2016) used the same dimension, but they called it completeness [3], [4]. In addition, some authors use the term quality dimension, some quality characteristic, and some even use the term quality metrics. Sadiq and Indulska (2017) pointed out insufficient examination of open data quality and have adduced that attention should be focused on three research areas in order to bring value to open information society: a shared, generally accepted understanding of the quality dimensions; the existence of awareness about the inherent quality of the data itself which means that collected data may be of sufficient quality for some purpose but not of sufficient quality for some other; examination of the relationship between data quality, intention to use such data and effective use of data in academic literature [5]. Open data portals should be observed independently of the software framework (e.g. CKAN, Socrata, OpenDataSoft etc.) in order to develop a quality assessment framework that can be applied to a number of portals rather than a few. This would provide a more complete picture of the current state of open data quality. By initial literature review another drawback in several researches was noted and it is about objectivity of developed quality frameworks. The objectivity of quality assessment frameworks is questionable since questionnaires were used for development as well as for evaluation of the same frameworks [1], [4]. Also, only the results of the surveys or evaluation of experts were used in a several phases during the development of the framework [1], [4].

The proposed research is motivated by existing frameworks in which the evaluation of open data quality is done. Even if results of the quality assessment in existing frameworks show a satisfactory level of portal quality, the quality of datasets published on them remains questionable. Therefore, open data portals that are rated as portals of good quality or which are ranked as the best can comprise datasets of poor quality.

Since further growth and development of open data is not possible assuming their questionable quality, the objective of the proposed research is to develop a framework for automatic evaluation of the quality of open datasets.

2. Research Methods

Considering the identified shortcomings in previous researches, the research questions and objectives are formulated. Combining quantitative and qualitative research approaches in social sciences is increasingly commonplace [6], [7]. Therefore, the research approach of proposed research is mixed methods because it consists of elements of qualitative and quantitative approaches [8]. The epistemological approach of the proposed research is pragmatism because it involves quantitative and qualitative methods, objective and subjective point of view, and acceptations of explanation that leads to the best outcomes [6], [9]. A method framework for design science research is used within proposed research along with modification in number of main activities [8].

Table 1. Plan of the proposed research

Activities	Research questions	Research objectives	Methods
Explicate Problem	What are the main quality dimensions of open datasets?	To synthesize the results of previous researches on the topic quality of open datasets and dimensions identified for the purpose of measuring them.	Systematic Literature Review
Define Requirements along with Design and Develop Artefact	How to measure identified quality dimensions of open datasets?	To develop a framework which is based on quality metrics of open datasets.	Expert Judgment Factor analysis Normalisation method
Demonstrate Artefact	How to apply a developed framework based on quality metrics of open datasets?	To apply developed quality framework to datasets available on open data portals.	Weights based on factor analysis Linear aggregation
Evaluate Artefact	How to evaluate results of demonstration of a developed quality framework of open datasets?	To evaluate the results of demonstration of a developed quality framework of open datasets.	Uncertainty analysis Sensitivity analysis

Dimension represents an aspect of the data quality which should be observed during the evaluation of open data [2], [4], [10].
Metrics represents the way of measuring open data quality and it partially or completely describes the quality of a dimension [2]–[4].

3. Results and discussion

Explicate Problem
Issues related to the quality of open datasets and their quality assessment are recognized by initial literature review. Further, the Systematic Literature Review will be used for the process of collecting, checking, and analysing existing literature on the quality of datasets and dimensions identified for the purpose of measuring them [11].
Define Requirements along with Design and Develop Artefact
Expert Judgment along with literature analysis will be used for the development of theoretical framework as well as for data selection [12], [13]. Theoretical framework is the step in the construction of the composite index and it is explained as the base which should be developed in such a way to provide a basis for selecting and combining individual indicators into a meaningful composite index that meets the principle of suitability for purpose. The step selection data implies checking the quality of the metrics, discussion of strengths and weaknesses of selected metrics and so on.
Various analytical approaches can be used to investigate whether the identified quality dimensions of the phenomenon are statistically well-balanced in the composite index [14]. So, factory analysis will be used in this step. The next step after multivariate analysis is normalization and it is necessary before any data aggregation since the indicators have different units of measurement in the dataset. At this phase of research, it could not be safely started which method will be chosen for normalization although there are a number of methods for normalisation of data (e.g. Ranking, Standardisation, Min-Max etc.) [13].
Demonstrate Artefact
According to the methodology for constructing the composite index, indicators or rather metrics in the framework should be weighted and aggregated with respect to the developed theoretical framework in the previous activity. Since the factor analysis is used in previous activity, the weighting based on factor analysis will be used for determination of weights. Nardo et al. (2005) pointed out that weighting based on factor analysis is possible only if there is correlation between individual indicators because otherwise it is unlikely to share common factors [13]. Although, there are different aggregation methods and each has its advantages and disadvantage, the linear aggregation will be used within proposed research since it is widely used [13], [15].
Evaluate Artefact
Uncertainty analysis and Sensitivity analysis will be used for evaluation of results obtained in previous activity Demonstrate Artefact. The combination of these two analyses can help measure the robustness of the composite index rankings, increase index transparency, and identify datasets that are ranked better or worse with respect to

certain assumptions. Estimation of composite index robustness by combining uncertainty analysis and sensitivity analysis proved to be useful in practice [13], [16].

4. Conclusion

Initial literature review on open data quality has revealed shortcomings in existing researches. Therefore, the framework for automatic evaluation of the quality of open datasets that is primarily based on metadata will be developed. For the development of the framework itself, it is necessary to identify key quality dimensions as well as the corresponding metrics.

The following 4 scientific contributions of the proposed research are: 1) systematization and synthesis of existing knowledge in the domain of the open datasets quality and dimensions identified for the purpose of measuring them; 2) the developed framework based on the quality metrics of open datasets solves the explained problem considering the specified requirements; 3) the developed framework shows that there is a difference in the quality of open datasets; 4) the developed framework is robust.

5. References

[1] S. Kubler, J. Robert, S. Neumaier, J. Umbrich, and Y. Le Traon, 'Comparison of metadata quality in open data portals using the Analytic Hierarchy Process', *Government Information Quarterly*, vol. 35, no. 1, pp. 13–29, Jan. 2018.

[2] J. Umbrich, S. Neumaier, and A. Polleres, 'Quality Assessment and Evolution of Open Data Portals', in *2015 3rd International Conference on Future Internet of Things and Cloud*, 2015, pp. 404–411.

[3] S. Neumaier, J. Umbrich, and A. Polleres, 'Automated Quality Assessment of Metadata Across Open Data Portals', *J. Data and Information Quality*, vol. 8, no. 1, pp. 2:1–2:29, Oct. 2016.

[4] A. Vetrò, L. Canova, M. Torchiano, C. O. Minotas, R. Iemma, and F. Morando, 'Open data quality measurement framework: Definition and application to Open Government Data', *Government Information Quarterly*, vol. 33, no. 2, pp. 325 – 337, Apr. 2016.

[5] S. Sadiq and M. Indulska, 'Open data: Quality over quantity', *International Journal of Information Management*, vol. 37, no. 3, pp. 150–154, Jun. 2017.

[6] I. Sekol and I. Maurović, 'MIJEŠANJE KVANTITATIVNOG I KVALITATIVNOG ISTRAŽIVAČKOG PRISTUPA U DRUŠTVENIM ZNANOSTIMA – MIJEŠANJE METODA ILI METODOLOGIJA?', *Ljetopis socijalnog rada*, vol. 24, no. 1, pp. 7–32, Jul. 2017.

[7] J. W. Creswell, *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*, 4th edition. Thousand Oaks: SAGE Publications, Inc, 2014.

[8] P. Johannesson and E. Perjons, *An Introduction to Design Science*. Springer International Publishing, 2014.

[9] R. B. Johnson and A. J. Onwuegbuzie, 'Mixed methods research: A research paradigm whose time has come', *Educational Researcher*, vol. 33, no. 7, pp. 14–26, 2004.

[10] N. Veljković, S. Bogdanović-Dinić, and L. Stoimenov, 'Benchmarking open government: An open data perspective', *Government Information Quarterly*, vol. 31, no. 2, pp. 278–290, Apr. 2014.

[11] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, 'Systematic Literature Reviews in Software Engineering - A Systematic Literature Review', *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009.

[12] M. A. Meyer and J. M. Booker, *Eliciting and Analyzing Expert Judgment: A Practical Guide*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2001.

[13] M. Nardo, M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman, and E. Giovannini, *Handbook on Constructing Composite Indicators*. OECD Publishing, 2005.

[14] W. K. Härdle and L. Simar, *Applied Multivariate Statistical Analysis*, 4th ed. Berlin Heidelberg: Springer-Verlag, 2015.

[15] R. K. Singh, H. R. Murty, S. K. Gupta, and A. K. Dikshit, 'An overview of sustainability assessment methodologies', *Ecological Indicators*, vol. 15, no. 1, pp. 281–299, Apr. 2012.

[16] M. Saisana, A. Saltelli, and S. Tarantola, 'Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 168, no. 2, pp. 307–323, 2005.

Internet of Things – The Next Industrial Revolution

Mihael Radovan
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
mradovan@foi.hr

1. Introduction

With the 21-th century, The Internet of Things came as a new industrial revolution – sometime referred to as Industry 4.0. All kinds of electronic devices and other non-electronic items and equipment (buildings, homes, cars, clothes and medicaments) can be equipped with sensors and actuators, which can then be connected to the Internet and exchange data. And data is the new commodity of the 21-st century, a new generator of values in the economy. This paper will try to give a review of possible use of IoT in most important industries – energy industry, automotive and smart communities (cities, buildings, homes).

2. Research Methods

Interview, Literature overview, Internet research

3. Results and discussion

The First industrial revolution came in the 18-th century with a steam machine, and mechanised production. A century later, the Second used oil and electricity to run the machines and created mass production. After WWII, with digital computers came the Third industrial revolution, which started automation era – robots and computer-aided machines started replacing humans on the production lines. Today, at the beginning of 21-st century, based on Cyber-Physical Systems, occurs the Fourth industrial revolution – Industry 4.0.
With omnipresent high-speed broadband internet access, 4-th generation mobile networks, cloud based internet services such a PaaS (Platform as a Service), Big Data and Machine-Learning, all new technologies developed in last years, it has become possible to connect almost any "thing" to the internet. Not just computers, but all other home devices and industrial machines, buildings, cars, roads, parking lots, agriculture, clothes, sport and medical devices – all these things can now become smart and start to send, receive and mutually exchange information.
Different kind of such "things", wireless interconnected, collect the data from sensors and send them to the data hub or cloud, and at the same time receive data from other devices or central systems, and coordinate its own behaviour in this way. Such kind of devices are becoming smaller and smaller, cheaper, more secure and powerful.
According to the statistical portal Statista (see Fig. 1), the number of connected "things" in 2017 is about 20 billion, forecast for 2020 is more than 30 billion devices, and for 2025 is more than 75 billion [1].

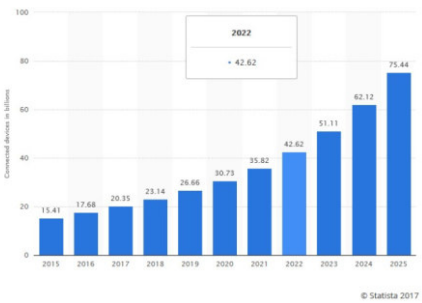

Figure 1. Number of connected IoT devices worldwide from 2015 to 2025

In general, IoT generates enormous amounts of data – this amount by 2020 will reach 600ZB per year – 4 times more

than in 2015. Most of these data will be ephemeral and will be neither saved nor stored. Approximately 10 percent of so generated data is useful and will be transmitted. Big Data will reach about 250 ZB, up about 10-fold from 2015, and global consumer cloud storage traffic from 8 EB in 2015 grows up to 50 EB in 2020, what is 6-fold growth [2].

4. IoT and Industry 4.0

Although most people under the term Internet of Things automatically think on consumer usage (smart home devices, sport gadgets or smart clothes), the Industrial Internet of Things (IIoT) began to capture significant attention in the world of business, and became lead technology for digital transformation and the No. 1 for 92% techno-companies.
In a last few years, IoT specific roles went "off the radar" in budgets and staffing plans of many big companies, who are important clients of Gartner. But at the beginning of 2016, 16% companies were concerned about IoT in their plans (see Fig. 2). This is a proof of rapidly growth of needs for IoT skills by the greater player in industry and business [4].

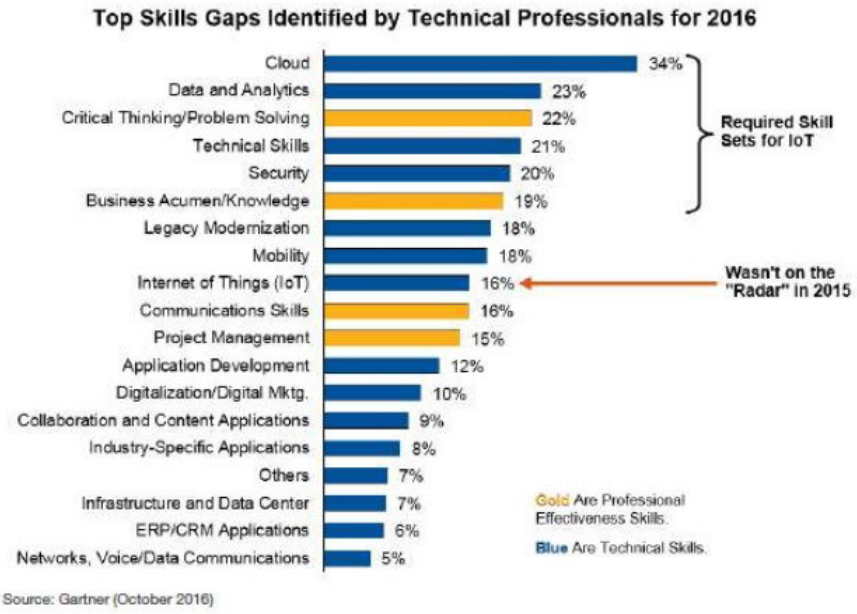

Fig. 2. Top industrial skills in 2016 by Gartner

How important IoT became, proves the fact that the biggest IT players offer IoT services and platforms to use: Microsoft, Oracle, Amazon, Cisco, Intel, Hitachi, Samsung, SAP, Google, Autodesk, IBM, Bosch, Siemens, Deutsche Telecom, AT&T, General Electric, Salesforce, but there are also many open source services.
The cloud has changed the way of software architecture and development of software for mobile platforms, which started to exchange petabytes of data, and finally parse, combine and analyse these data in the cloud, producing a new values and powerful insights to business. The amount of data generated by IoT continues to grow at an exponential rate. IoT is much more than technology, much more than devices connected to the cloud. IoT is a new techno-philosophy, it is a way of thinking. It is a fantastic opportunity for industry to enter new markets and develop new business models and new industrial standards. Creating new business opportunities is enough big reason companies are looking to the IoT today. In most cases, the purpose of business is to increase profit.
Every new technology always and absolutely meets a child diseases and many treatment. By IoT, this are inappropriate existing industrial network protocols, as well as security models. Security is the most important part of such a system, from physical protecting of devices at the edge to protecting customer's data and privacy in the cloud. IoT security is as strong as the weakest node. Therefore is by IoT used a new technology, approved by Bitcoin – Blockchain. "Blockchain has relevance to IoT in establishing the provenance of devices in a specific network; in particular where long running but infrequently connected remote devices can report in", says Ian Hughes, IoT analyst at 451[1] Research in London.

[1] Ian Hughes, 451 Research, London, UK: https://451research.com/analyst-team/analyst/Ian+Hughes

One of biggest challenges for IoT, outside of industry, and big business opportunity are smart cities. IoT slowly and quietly enters in our homes [4] – we already have smartphones, smart lightning, smart TVs, smart heating and cooling devices, connected security systems, smart locks, WLAN routers with support for smart home devices. Next step are smart buildings [5] and finally smart cities [3], where is almost every connected to the internet – smart buildings, connected cars, smart parking lots, connected public traffic and transportation, security personal etc. One big problem and challenge in this case is how to handle with such a big amount of data. Just one smart city with the population of one million residents will by 2020 generate 200 PB (200 million GB) of data per day (see Table 1.)

Table 1. Generated and transmitted data in the Smart City

Thing	Generated	Transmitted	% Transmitted
Weather Station	10 MB	500 KB	5%
Smart Grid	5 GB	50 MB	1%
Smart Building	275 GB	2,75 GB	1%
Smart Car	70 GB	70 MB	0,10%
Smart Hospital	5 TB	5 GB	0,10%
Connected Plane	40 TB	40 GB	0,10%
Smart Factory	1 PB	2 TB	0,20%
Public Safety	50 PB	50 TB	0,10%

Despite security threats and problems with infrastructure and communication, IoT is a solution for traffic problems of megacities. According Inrix Research [7], car drivers spent 9% their time sitting in traffic last year. For this study, Inrix Research covered 100.000 locations across 8700 cities in more than 100 countries, and about 18.000 drivers. To reduce enormous costs for the parking overpaid and time spent for search for parking, it is necessary to interconnect cars with the city infrastructure, using IoT technology – connected cars and smart parking solutions can be leveraged to reduce searching for parking, and save time and fuel.

5. Conclusion

About half of the world's population lives in cities. There are more than 20 megacities of 10 million or even more people, and in next 10 years this amount will be doubled. Technological development must be used to manage with resources on optimal way. Smart cities and smart industry solutions are keys to have a sustainable development in this important environments.

6. References

[1] Statista, "Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025," 2017. Available: https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/.

[2] Cisco, "Cisco Global Cloud Index: Forecast and Methodology, 2015 - 2020," 2016. Available: https://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11738085.pdf.

[3] Gartner, "2017 Planning Guide for the Internet of Things," 2017. Available: http://images.gartnerformarketers.com/Web/Gartner/%7B0a313d84-f153-42fe-a6319d4860eea824%7D2017_planning_guide_for_the__IoT.pdf

[4] D. Geneiatakis, I. Kounelis, R. Neisse, I. Nai-Fovino, G. Steri i G. Baldini, »Security and Privacy Issues for an IoT based Smart Home,« u Mipro, Opatija, Croatia, 2017.

[5] S. H. Maciel, I. Cardoso, H. S. Ramos, J. P. C. Rodrigues and A. L. L. Aquino, "An integrated access control and lighting configuration system for smart buildings," Journal of communications software and systems, vol. 13, pp. 101-108, 06/2017.

[6] Inmarsat, "The Future of IoT in Enterprise - 2017," 2017. [Online]. Available: https://www.inmarsat.com/wpcontent/uploads/2017/06/IRP The_Future_of_IoT_in_Enterprise_2017.pdf.

[7] G. P. B. Cookson, »The Impact of Parking Pain in the US, UK and Germany,« Inrix Research, 2017.

Process Parameters Discovery Based on Application of k-Means Algorithm - A Real Case Experimental Study

Snježana Križanić, Katarina Tomičić-Pupek
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
{skrizanic, katarina.tomicic}@foi.hr

Abstract

This paper describes the application of the k-means algorithm on educational data to detect processes in the e-learning environment at a Higher Education institution in Croatia. Based on the interpretation of analysed data sets, two illustrative examples of process models are given, showing time reference and activity frequency as valuable process parameters extractable from raw data. The goal of this paper is to explore the application of a grouping method over educational logs with the aim to identify potential process parameters needed to understand student behaviour in accessing teaching material.

Keywords: process discovery, process parameters, k-means, educational event logs

References

[8] Bizagi, **Digital Process Automation and BPM**, available at https://www.bizagi.com/, Accessed: 19th July 2019.

[9] Bose, R.P.J.C., Van der Aalst, W.M.P., **Trace clustering based on conserved patterns: Towards achieving better process models,** Lecture Notes in Business Information Processing, vol. 43, 2010, pp. 170-181.

[10] Buijs, J. C. A. M., Van Dongen, B. F., Van der Aalst, W. M. P., **Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity**, International Journal of Cooperative Information Systems, 23(1), 2014.

[11] Campagni, R., Merlini, D., Sprugnoli, R., Verri, M.C., **Data mining models for student careers**, Expert Systems with Applications, 42(13), 2015, pp. 5508-5521.

[12] De Leoni, M., Van der Aalst, W. M. P., Dees, M., **A General Process Mining Framework for Correlating, Predicting and Clustering Dynamic Behavior Based on Event Logs**, Information Systems, 56, 2016, pp. 235-257.

[13] Dumas, M., La Rosa, M., Mendling, J., Reijers, H. A., **Fundamentals of Business Process Management**, Springer-Verlag, Berlin, Heidelberg, 2013.

[14] Dutt, A., Ismail, M.A., Herawan, T., **A Systematic Review on Educational Data Mining**, IEEE Access, 5, 2017, pp. 15991-16005.

[15] García-Bañuelos, L., Dumas, M., La Rosa, M., De Weerdt, J., Ekanayake, C.C., **Controlled automated discovery of collections of business process models**, Information Systems, 46, 2014, pp. 85-101.

[16] RapidMiner Studio, **Visual workflow designer for the entire analytics team**, available at https://rapidminer.com/products/studio, Accessed: 19th July 2019.

[17] Sarno, R., Effendi Y. A., Haryadita, F., **Modified Time-Based Heuristic Miner for Parallel Business Processes**, International Review on Computers and Software, 11(3), 2016, pp. 248-260.

[18] Van der Aalst, W., Adriansyah, A., Van Dongen, B., **Replaying history on process models for conformance checking and performance analysis**, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(2), 2012, pp. 182-192.

Classification of Motor Imagery Tasks for Brain-Computer Interface with SVM Classifiers

Marcel Nikmon, Roman Budjač

Faculty of Materials Science and Technology in Trnava / Institute of Applied Informatics, Automation and Mechatronics

Slovak University of Technology in Bratislava

J. Bottu 25, 917 24 Trnava, Slovakia

{marcel.nikmon, roman.budjac}@stuba.sk

1. Introduction

Brain - computer interface is created to control devices without physical movement. This type of interfaces works on the principle of acquisition of information from the user's brain by measuring brain activity. To measure brain activity is used EEG method, which captures the electric potentials between the neurons using non-invasive electrodes located on the user's skull. Our ongoing research is focused on looking for thought patterns that will be used to drive machines or equipment, which routine control is not possible with physical movement, mainly because of the user's inability to perform physical movement that may be caused by a nervous system disease or injury. For this purpose is used motor imagery, that research suggests, that thought movement is very similar to real move EEG data. [1] The aim of the research is to find thought patterns that can be detect during execution, not just when it is activated, as is the case with P300 waves, which are generated 300ms after stimulus or after real/imagery movement is intended. In this paper, we focus on the classification of ongoing movement in 3 categories, using the SVM machine learning methods with various parameters provided by Matlab.[2]

2. Research Methods

The aim of the research is to find a thought pattern that can be captured throughout its execution and can be safely distinguished from other thoughts that are natural to the human brain, such as breathing, blinking, swallowing, etc. At the same time, such an thought cannot be hard to focus for the user, as he would be mentally very exhausted in long-term handling of BCI. For our purposes, we have collected data for only a small time period in order to train classifier in a relatively short period of time. The results of this research will tend as benchmarks for the future design of a neural network designed for the same purpose.

3. Data acquisition

We use the data acquired by the EEG headset Emotiv Epoc+ to input SVM classifiers. This device is capable of recording 16-bit electrical potentials up to 8400 μV with 128Hz. The location of the electrodes is in the standard international 10-20 system. The recorded bandwidth is 0.16Hz - 43Hz. Data is recorded using OpenVibe software, which is OpenSource licensed. [3]

4. Data processing

After we have acquired raw EEG data, we created an application that processes this data. This application speededup the data editing and thus the whole training process. Processing is carried out in several steps. The first step is to create a label for each data category. In our case, they are 3 states: neutral/calm, imagery left hand movement, and imagery right hand movement. From these data, we cut 500ms at the beginning and end of each recorded data class to avoid detecting the P300 waves that arise at the beginning of the intended movement.

5. Data filtering

Data is filtered using bandpass filter in the following ranges. Alpha - 7Hz-13Hz, Beta 14Hz-30Hz and Epoc device range 0.16Hz-43Hz. From each frequency band, we created one training dataset. For our purposes, we have also created one combined dataset, which contain Alpha + beta waves in the range of 7-30.

6. FFT filtering

Next we applied fast fourier transform to data processed by bandpass filter. This transformation changed the data from the time domain to the frequency domain and then we tried training on the data processed in this way. Since FFT requires 2^N processing mask, we had to choose the appropriate mask size to be applied to the data. We chose 128, that is 2^7, that represent data for 1 second. This can be relatively fast response to the classification of the thought pattern. If we chose a larger mask, the classifier responses would be slow, so the control signal could be executed after more than 1 second. This is no longer considered to be quick enough for real time applications.

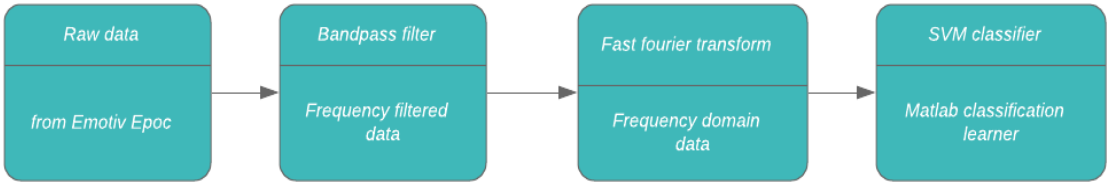

Figure 1. Data processing

7. Classifiers

As mentioned above, we use support vector machine classifiers for classification. Specifically, they were SVM quadtratic, SVM cubic and SVM fine gaussian. The training dataset was created for 3 classes: lefthand imagery movement, right hand imagery movement and neutral state. We used cross validation method with 5 crossfolds for data validation. This protects classifier against overfitting by partitioning the dataset into folds and estimating accuracy on each fold. [4]

8. Results and discussion

During the first trainings of the SVM classifier, the results were very close to the random distribution, ie. at 3 classification classes that was about 33% - 40% accuracy of classification. These results were mainly in linear SVM but also in Cubic and Quadratic SVM. For SVM Gaussian, the results were in the range of 46% -66%, which still did not provide sufficient accuracy.

Table 1. Classification accuracy

Frequency range	SVM quadratic	SVM cubic	SVM fine gaussian
0.16 Hz – 43 Hz (default Epoc device range)	33,7%	33,3%	95,0%
7 Hz – 13 Hz (alfa)	33,4%	33,9%	93,2%
14 Hz -30 Hz (beta)	33,3%	33,7%	95,7%
7 Hz -30 Hz (alfa+beta)	33,6%	33,3%	95,0%

Changing the kernel scale training parameter to the automatic mode provided by the Matlab software when training SVM gaussian type has shown us very nice results, see Table 1. Classification accuracy. Detection accuracy increased to more than 93%. Best classification accuracy was in Beta waves trained by SVM fine gaussian method. That is shown in Figure 2. Confusion matrix shows that the highest error rate was in neutral state class and the most successful classification was in motor imagery classes. This is because the motor imagery tasks is similar to each other and very different from the neutral state.

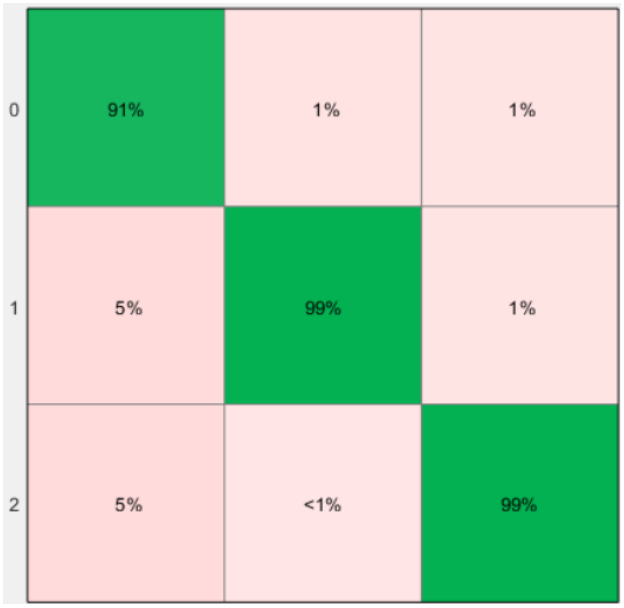

Figure 2. Confusion matrix

9. Conclusion

Research has shown that classifier parameters such as kernel scale are very important to the classifier accuracy. We proved that linear, quadratic and cubic SVM are not suitable for EEG data and that beta waves have the most significant results. These results were trained on small data sets. The next step that will be used to verify the use of such an SVM classifier in the real world will be to record new data with the intention of verifying the accuracy of the classification under real conditions. If the success of the classification proves to be sufficient, then we will record the data under difficult conditions such as noise, changing light conditions and other disturbing elements from the environment. If we found out that the classification is also highly successful, the implementation of the classifier into the real machine control algorithm would follow. This BCI could actually serve physically challenged people. This research originally served as a starting point for creating a BCI based on neural networks but shows that SVM methods can also be used on the motor imagery tasks classification.

10. Acknowledgments

This publication was created thanks to the support of the Ministry of Education, Science, Research and Sport of the Slovak Republic in the framework of the call for subsidy for the development project No. 002STU-2-1/2018 with the title „ STU as the Leader of the Digital Coalition“.

11. References

[1] Toderean Aldea, Roxana & Fira, Monica. (2014). Classification of motor imagery task in Brain computer interface using linear discriminant analysis. *Internal Journal of Avanced Research in Artificial intelligence.*

[2] Malihe S., Reza B., Separation of P300 event-related potential using time varying time-lag blind source separation algorithm. *Computer Methods and Programs in Biomedicine* 145 (2017) pages 95-102.

[3] Emotiv Epoc Specification, (acessed 1.6.2019 https://www.emotiv.com/product/emotiv-epoc-14-channel-mobile-eeg/#tab-description)

[4] Raju, Srujan & Murty, M. (2018). Support Vector Machine with K-fold Cross validation model for software fault prediction. *International journal of Pure and Applied Mathematics.* 118. 321-334.

Combining Sentiment Analysis with Linguistics: Figurative Speech Detection in Croatian Online Text

Martina Ptiček
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
marpticek@foi.hr

Abstract

Figurative speech (i.e. language figures such as irony, sarcasm, metaphor etc.) makes an important part of communication including the online communication and texts, such as posts, reviews, tweets etc. [1, 240]. In order to properly analyse the sentiment of texts, attention should be given to the fact that their meaning sometimes can be "hidden" due to the use of figurative language [1, 240], [2, 7].

Knowing that recognition of figurative language can be a challenge even for a human being, analysing and recognizing figurative language is an especially challenging task in natural language processing and in machine learning [1], [2], [3], [4]. The analysis has to be twofold and in first instance encompass decoding of the literal meaning after which the second code can be decoded that will lead to identification of the "real" meaning.

To develop a model of a sentiment analysis [5], [6], [7], [8] that will take into the account figurative language, this doctoral research will start from analysing irony, as the common figure in the use of figurative language. Irony will be analysed from the linguistical perspective with the goal to identify important aspects which will allow us to discriminate ironic from non-ironic sentence. After these characteristics are defined, a model [9], [10], [11] will be developed which will be able to discern between the ironic and non-ironic sentence.

Developed model will be tested in conducting the task of sentiment analysis this test will be used to support the hypothesis that sentiment analysis will be more successful if the model takes into account that figurative speech is present, compared to the sentiment analysis that does not recognize the figurative speech.

The dataset that will be used for this experiment will consist of user-generated content published on Croatian forums and comments on selected news portals. It is expected that this model won't give the same results in different domains (e.g. politics, sport, fashion etc.) [5], [6], [7], [11], [12] therefore its relevance will be tested for each of the selected domains, which will enable a comparative insight into the applicable value of the model in specific domains.

Keywords: web text analysis, figurative speech, figurative language processing, irony, sentiment analysis

References

[19] Antonio Reyes, Paolo Rosso, Tony Veale: **A Multidimensional approach for detecting irony on Twitter**, Language Resources and Evaluation, Berlin, Heidelberg, Germany, 2013, pp. 239-268

[20] Antonio Reyes, Paolo Rosso, Davide Buscaldi: **From humor recognition to irony detection: The figurative language of social media**, Data & Knowledge Engineering, Amsterdam, Netherland, 2012, pp. 1-12

[21] Aniruddha Ghosh, Tony Veale, Ekaterina Shutova, John Barnden, Guofu Li, Paolo Rosso, Antonio Reyes: **SemEval-2015 Task 11: Sentiment Analysis of Figurative language on Twitter**, Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 4th – 5th June, Denver, Colorado, 2015, pp. 470-478.

[22] Diana Maynard, Mark Greenwood: **Who cares about Sarcastic Tweets? Investigating the**

Impact of Sarcasm on Sentiment Analysis, Proceedings of the Ninth International Conference on Language Resources and Evaluation, May, Reykjavik, Iceland, 2014, pp. 4238–4243

[23] Bing Liu: **Sentiment Analysis**, Cambridge University Press, New York, NY, USA, 2015.

[24] Bo Pang, Lillian Lee: **Opinion mining and sentiment analysis**, Foundations and trends in Information retrival, Hanover, USA, 2008, pp. 1-135

[25] Maite Taboada, Julian Brooke, Milan Tafiloski, Kimberly Voll, Manfred Stede: **Lexicon-Based Methods for Sentiment Analysis**, Computational Linguistics, Cambridge, Massachusetts, 2013., pp. 267 - 307

[26] Daniel Jurafsky, James H. Martin: **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**, available at `https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf`, Accessed: 14th July 2019.

[27] Eugene Charniak: **Introduction to deep learning**, The MIT Press, Cambridge, Massachusetts, 2018.

[28] Thorsten Joachims: **Text Categorization with Suport Vector Machines: Learning with Many Relevant Features,** Machine Learning: ECML-98. ECML 1998. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol 1398. Springer, Berlin, Heidelberg, pp. 137-142

[29] Saif Mohammad, Peter D. Turney: **Crowdsourcing the Creation of a Word-Emotion Association Lexicon**, Computational Intelligence, New Jersey, 2013, pp. 436-465

[30] William L. Hamilton, Kevin Clark, Jure Leskovac, Dan Jurafsky: **Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora**, Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas, 2016, pp. 595 – 605

Sentiment Analysis of Text Documents

Dalibor Bužić
College for Information Technologies
Klaićeva 7, Zagreb, Croatia
dalibor.buzic@vsite.hr

Abstract

By popularizing Web 2.0 technologies, the internet has become a rich source of user-generated content. Due to the large amount of user content, a need for its efficient computer processing emerged very quickly. Sentiment analysis, one of the areas of application within the natural language processing, collects and processes people's opinions and attitudes toward products, services, politics, social events, marketing campaigns and company strategies.

In the sentiment analysis, unstructured text is processed, and it brings numerous problems in computer processing. Some of the obstacles to successful analysis are the distinction between fact and opinion, syntax errors, negations, comparative sentences, irony and sarcasm recognition, fake opinions, context dependency, domain dependency, etc. In order to overcome problems, different methods and techniques are used.

This paper gives a literature overview in the field of sentiment analysis focusing on current research approaches. For a successful sentiment analysis, proper preprocessing and feature engineering is required. Then the application of the techniques and methods of one of the two fundamental approaches (lexicon-based and machine learning) to the sentiment analysis follows. The lexicon-based approach has two variants: using a dictionary or a corpus-based approach. Depending on whether there are annotated learning data sets or not, machine learning can be supervised or unsupervised.

Dataless classification is a learning protocol that uses general knowledge to train the classifier without the need for any labeled data, and it can be applied in sentiment analysis.

Deep learning, an approach that uses artificial neural networks in learning tasks through the network of multiple layers, has recently become increasingly popular in sentiment analysis.

A relatively new subfield within sentiment analysis is emotion mining. Instead of recognizing positivity or negativity, researchers are increasingly investigating recognition of concrete emotions in the text. Emotion mining is in the early phase of the research community, and it suffers from a lack of labeled data (great effort has to be done in text tagging). The successful application of emotion mining could stimulate more interest in research in this direction.

Keywords: sentiment analysis, opinion mining, machine learning, deep learning, dataless classification, emotion mining

Acknowledgments

This poster is part of final paper "Bužić, D: Sentiment analysis of text documents, CECIIS 2019, Varaždin, Croatia".

References

[1] Agrawal, A., An, A., Papagelis, M.: **Learning Emotion-enriched Word Representations**, Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 950-961.

[2] Al-Moslmi, T., Omar, N., Abdullah, S., Albared, M.: **Approaches to cross-domain sentiment analysis: A systematic literature review**, IEEE Access, 5, 2017, pp. 16173-16192.

[3] Appel, O. et al.: **Successes and challenges in developing a hybrid approach to sentiment analysis**, Applied Intelligence, 48(5), 2018, pp. 1176-1188.

[4] Badjatiya, P. et al.: **Deep learning for hate speech detection in tweets**, Proceedings of the 26th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, 2017, pp. 759-760.

[5] Bravo-Marquez, F., Mendoza, M., Poblete, B.: **Combining strengths, emotions and polarities for boosting Twitter sentiment analysis**,

Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining (p. 2), ACM, 2013.

[6] Chang, M.W. et al.: **Importance of Semantic Representation: Dataless Classification**, Proceedings of the 23rd AAAI conference on Artificial intelligence, Vol. 2, 2008, pp. 830-835.

[7] Chapman, L. et al.: **Investigating the emotional responses of individuals to urban green space using twitter data: A critical comparison of three different methods of sentiment analysis**, Urban Planning, 3(1), 2018, pp. 21-33.

[8] Chaturvedi, I. et al.: **Distinguishing between facts and opinions for sentiment analysis: Survey and challenges**, Information Fusion, 44, 2018, pp. 65-77.

[9] Chen, X. et al.: **Dataless text classification with descriptive LDA**, Proceedings of the 29th AAAI conference on Artificial intelligence, 2015, pp. 2224-2231.

[10] Choo, E., Yu, T., Chi, M.: **Detecting opinion spammer groups and spam targets through community discovery and sentiment analysis**, Journal of Computer Security, 25.3, 2017, pp. 283-318.

[11] Esuli, A., Sebastiani, F.: **SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining**, 5th Conference on Language Resources and Evaluation, 2006, pp. 417-422.

[12] Goldberg, Y.: **A primer on neural network models for natural language processing**, Journal of Artificial Intelligence Research, 57, 2016, pp. 345-420.

[13] Jindal, N., Liu, B.: **Opinion spam and analysis**, Proceedings of the 2008 international conference on web search and data mining, ACM, 2008, pp. 219-230.

[14] Liu, B.: **Sentiment Analysis and Subjectivity**, Handbook of natural language processing, 2, 2010, pp. 627-666.

[15] Liu, B.: **Sentiment analysis: Mining opinions, sentiments, and emotions**, Cambridge University Press, New York, 2015.

[16] Madhoushi, Z., Hamdan, A.R., Zainudin, S.: **Sentiment analysis techniques in recent works**, Science and Information Conference (SAI), IEEE, 2015.

[17] Martins, R. et al.: **Predicting Performance Problems Through Emotional Analysis (Short Paper)**, 7th Symposium on Languages, Applications and Technologies (SLATE 2018), Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[18] Maynard, D., Greenwood, M.A.: **Who cares about Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis**, Lrec, 2014, pp. 4238-4243.

[19] Medhat, W., Hassan, A., Korashy: **Sentiment analysis algorithms and applications: A survey**, Ain Shams Engineering Journal, 5.4, 2014, pp. 1093-1113.

[20] Mikolov, T. et al.: **Efficient estimation of word representations in vector space**, arXiv preprint arXiv:1301.3781, 2013.

[21] Mohammad, S. M.: **Sentiment analysis: Detecting valence, emotions, and other affectual states from text**, Emotion measurement, 2016, pp. 201-237.

[22] Mohammad, S. M., Kiritchenko, S.: **Using nuances of emotion to identify personality**, Proceedings of ICWSM, 2013, pp. 27-30.

[23] Paltoglou, G., Thelwall, M.: **More than bag-of-words: Sentence-based document representation for sentiment analysis**, Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013, 2013, pp. 546-552.

[24] Pang, B., Lee, L.: **Opinion mining and sentiment analysis**, Foundations and Trends® in Information Retrieval, 2(1–2), 2008, pp. 1-135.

[25] Pennington, J., Socher, R., Manning, C.: **Glove: Global vectors for word representation**, Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532-1543.

[26] Plaza-del-Arco, F. M. et al.: **Lexicon Adaptation for Spanish Emotion Mining**, Procesamiento del Lenguaje Natural, 61, 2018, pp. 117-124.

[27] Ravi, K., Ravi, V.: **A survey on opinion mining and sentiment analysis: tasks, approaches and applications**, Knowledge-Based Systems, 89, 2015, pp. 14-46.

[28] Singh, T., Kumari, M.: **Role of text pre-processing in twitter sentiment analysis**,

Procedia Computer Science, 89, 2016, pp. 549-554.

[29] Smailović, J. et al.: **Predictive sentiment analysis of tweets: A stock market application**, Human-computer interaction and knowledge discovery in complex, unstructured, Big Data, Springer, Berlin, Heidelberg, 2013, pp. 77-88.

[30] Song, Y., Roth, D.: On Dataless Hierarchical Text Classification, Proceedings of the 28th AAAI conference on Artificial intelligence, Vol. 7, 2014, pp. 1579-1585.

[31] Vadicamo, L. et al.: **Cross-Media Learning for Image Sentiment Analysis in the Wild**, ICCV Workshops, 2017, pp. 308-317.

[32] Xu, C. et al.: **Visual sentiment prediction with deep convolutional neural networks**, arXiv preprint, arXiv:1411.5731, 2014.

[33] Zhang, L., Wang, S., Liu, B.: **Deep learning for sentiment analysis: A survey**, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253, 2018.

[34] Zhang, Y. et al.: **iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization**, Future Generation Computer Systems, 66, 2017, pp. 30-35.

Evaluation of the Voice to Text Transfer in Augmented Conditions

Matej Janíček, Karol Velíšek, Radovan Holubek, Roman Ružarovský
Faculty of Materials Science and Technology in Trnava
Slovak University of Technology in Bratislava
Jána Bottu 2781/25, 917 24 Trnava, Slovakia
{matej.janicek, karol.velisek, radova.holubek, roman.ruzarovsky}@stuba.sk

Abstract

The aim of the research was evaluation of voice to text transfer in augmented conditions for the use in the verbal control of industrial robots. Research builds on previous research, evaluation of voice to text transfer in different conditions, where the main problem of the voice to text transfer has been defined. It was clear from the results that the main problem of voice to text transfer is ambient noise. In this research, a device for eliminating ambient noise was used – the limiter. Simulation was designed and implemented to eliminate the main problem of voice to text transmission, the same procedure of simulation were used. The results of previous research were compared with the results of current research. Based on the analyzes, the result of this research was determined.

Keywords: voice to text transfer, verbal control, evaluation

Acknowledgments

This publication was created thanks to the support of the Ministry of Education, Science, Research and Sport of the Slovak Republic in the framework of the call for subsidy for the development project No. 002STU-2-1/2018 with the title „ STU as the Leader of the Digital Coalition.

References

[1] Windmann, S., & Haeb-Umbach, R. (2009). Approaches to Iterative Speech Feature Enhancement and Recognition, **IEEE Transactions On Audio, Speech, And Language Processing**, Vol. 17, No. 5.

[2] Gustavssona, P., Syberfeldta, A., Brewsterb, R. & Wangc, L. (2017). **The 50th CIRP Conference on Manufacturing Systems**, Procedia CIRP, 63, 396 – 401.

[3] Rogowski, A. (2012). Industrially oriented voice control system, **Robot. Comput. Integr. Manuf.,** 28, 303–315.

[4] Kohanski, M., Lipski, A., M., Tannir, J. & Yeung, T. (2002**). Development of a Voice Recognition Program**, available at www.seas.upenn.edu/~belab/LabProjects/2001/be310s01t2.doc, Accessed: 15th September 2019.

[5] Rogowski, A. (2013). Web-based remote voice control of robotized cells. **Robot. Comput. Integr. Manuf.**, 29, 77–89.

[6] Gundogdu, K., Bayrakdar, S. & Yucedag, I. (2018) **Journal of King Saud University – Computer and Information Sciences**, 30, 198–205.

[7] Qadri, M. & Ahmed, S.A. (2009). **IEEE International Conference on Signal Acquisition and Processing**, 217– 220.

[8] Jayasekara, B., Watanabe, K. & Izumi, K. (2008). **SICE Annual Conference**, 1, 2540–2544.

Methodological framework for efficient prediction of carbon dioxide emissions in building sector by intelligent data analytics as fundamental module of an integrated energy management information system

Saša Mitrović
PhD student at Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varazdin, Croatia
smitrovic@foi.hr

1. Introduction

The importance of increasing energy efficiency in building sector was emphasized by the European Council in Decisions: 2010/31/EU and 2012/27/EU, which stated that buildings account for 40% of total energy consumption in the European Union, and that the building sector is expanding, which will increase energy consumption in the future. Thus, due to the need to increase energy efficiency and reduce greenhouse gas emissions, European Union has, through the aforementioned decisions, set a target of 20% reduction in energy consumption by 2020. [1, 2] In December 2018, the revised Energy Efficiency Decision 2018/2002 enters into force, which sets a new EU energy efficiency target - to reduce energy consumption by at least 32.5% by 2030. [3]. In addition to the aforementioned decisions, Decision 406/2009/EC established binding targets for State Members to reduce carbon dioxide (hereinafter referred to as: CO2) by year 2020 by 20% and by year 2030 by 40%. Energy efficiency in the building sector is crucial to meeting the binding targets, and in order to achieve efficiency it is necessary to reduce the energy consumption of buildings in the total energy consumption of all sectors [4].

Tommerup, et al. [5] states that the building sector with 40% of total energy consumption compared to the sectors of industry and transport is the largest single energy consumer in the European Union [5], while the building sector in Croatia accounts for 30% of total energy consumption [6]. Aforementioned percentages are an indicator that energy efficiency in buildings should also be establish in Croatia in order to reduce CO2 emissions. Croatia, in order to ensure energy efficiency, followed guidelines of the European Union decisions and passed the Energy Efficiency Act in the Croatian Parliament [7]. In addition, representatives of all institutions are invited to apply the ISO 50001 standard, which sets out the requirements for establishing an energy management system in the building sector [8].

The energy management information system in the building sector should ensure energy efficiency to reduce CO2 emissions. Such an information system should predict CO2 emissions in the future based on past data. Machine learning methods and models or mathematical methods and models can be used for such forecast. Authors Wang and Tao [9] indicate the following conclusions „The two fundamental problems in machine learning (ML) are statistical analysis and algorithm design. The former tells us the principles of the mathematical models that we establish from the observation data. The latter defines the conditions on which implementation of data models and data sets rely.“[9] According to the author's Wang and Tao [9] it can be concluded that mathematical models only describe observed data without the ability to learn on past data, whereas machine learning models provide the ability to learn from observed data [9] so the author of this paper presume that it is better to apply machine learning models over only mathematical models, hence, only machine learning will be presented down from point of paper. Burkov [10] defined machine learning as „the process of solving a practical problem by 1) gathering a dataset, and 2) algorithmically building a statistical model based on that dataset“[10]. One of the essential methods of machine learning according to Jain [11] is cluster analysis, which is defined as „formal study of methods and algorithms for grouping, or clustering, objects according to measured or perceived intrinsic characteristics or similarity“[11] and for which he emphasized the following advantage: „Organizing data into sensible groupings is one of the most fundamental modes of understanding and learning“[11]. For author's Jordan and Mitchell [12] machine learning „addresses the question of how to build computers that improve automatically through experience.“[12] Jordan and Mitchell [12] for the limitations and challenges of machine learning stated the following: „most machine

learning algorithms are targeted to learn one specific function or data model from one single data source, humans clearly learn many different skills and types of knowledge, from years of diverse training experience, supervised and unsupervised, in a simple-to-more-difficult sequence (e.g., learning to crawl, then walk, then run). "[12] Authors Jordan and Mitchell [12] identified in their research paper of machine learning methods that lately supervised learning machine learning methods have been used mostly with application in spam detection, facial recognition of images and in diagnostic systems in medicine where they have achieved higher rates of accurate prediction with respect to other machine learning methods [12]. Supervised learning, defined by Shobha and Rangaswamy [13], is „a learning model built to make prediction, given an unforeseen input instance. A supervised learning algorithm takes a known set of input dataset and its known responses to the data (output) to learn the regression/classification model. A learning algorithm then trains a model to generate a prediction for the response to new data or the test dataset."[13] Shobha and Rangaswamy [13] stated that „Classification task predicts discrete responses. It is recommended if the data can be categorized, tagged, or separated into specific groups or classes."[13] According to Talabis, et al. [14] supervised learning methods are „focused on the prediction of qualitative responses by analysing the input dataset and identifying patterns." [14] McCue [15] stated that the aim of supervised learning methods are to „develop a set of decision rules used to determine outcomes."[15] Due to the explanations and reasons stated by author's Jain [11], Jordan and Mitchell [12], Shobha and Rangaswamy [13], Talabis, et al. [14] and McCue [15], author of this research paper presume that it is worthy to use machine learning methods, in particular clustering and supervised learning methods, to create a methodological framework for reliable prediction of CO2 emissions in building sector. According to the aforementioned authors, clustering parameters needs to be well prepared in advance, also the number of clusters should be determined in advance, which is a disadvantage of this method, but when the data is well prepared for processing and a good number of clusters are identified, the clustering algorithm produces valuable results. For the supervised learning methods, aforementioned authors stated that the models should be well adjusted before training on the data and that larger amounts of training data are needed, which is a disadvantage, but not in the data of this paper as it covers a 10-year period, while supervised learning methods are good at creating rules and nonlinear decision boundaries if the models are well adjusted.[11-15]

Thus, the aim of this research paper is to develop a methodological framework using intelligent data analytics with a cluster procedure and an optimal architecture of machine, supervised, learning methods that can process large amounts of data to predict CO2 emissions in buildings on an annual basis in Croatia as an integral part of the information system for energy management with the aim of reducing CO2 emissions and meeting the binding targets of the Republic of Croatia as a State Member of the European Union for the reduction of CO2 emissions.

According to the Dresner Advisory Services [16] study from 2019 data analytics, machine learning and other advanced algorithms are ranked as the eighth priority among the 37 technologies used by companies to develop and achieve business plans and goals [16]. Given that the eighth place of priorities for the development of artificial intelligence and machine learning models in companies shows, according to the aforementioned research, the importance of using data analytics in achieving business plans, therefore, the use of artificial intelligence and machine learning models in developing a methodological framework for predicting CO2 emissions should be amidst the top priorities among the various CO2 emission monitoring technologies.

Numerous studies have been carried out to predict CO2 emissions and energy consumption using different machine learning methods and algorithms. Hong, et al. [17] developed a model for predicting national CO2 emissions in buildings for 2030 using evolutionary algorithms: gene expression programming - harmony search algorithm and the developed model had mean absolute percentage error (hereinafter referred to as: MAPE) 2.06% [17]. Sun and Liu [18] used a support-vector machines method to predict CO2 emissions which resulted in a high rate of accurate prediction - 0.160% (residential buildings) and 0.328% (three major industries and residential buildings) MAPE [18]. Kankal, et al. [19] have proposed artificial neural networks to predict energy consumption in the Turkish building sector [19]. Jeong, et al. [20] used data mining methods (correlation analysis, decision trees, Analysis of variance (ANOVA)) and Monte Carlo simulations to predict energy use and CO2 emissions to build energy efficient model of residential buildings. Kontokosta and Tull [21], Melo, et al. [22], Son, et al. [23], Farzana, et al. [24] and Chou and Bui [25] also, among various methods, used machine learning methods: artificial neural networks, support-vector machines, support-vector regression to predict energy consumption in residential, commercial and other public sector owned buildings [21-25]. Naji, et al. [26] in Naji, et al. [27] used support-vector regression method and adaptive network-based fuzzy inference system (ANFIS) to predict the energy consumption of buildings made in the Energy Plus building energy simulation program [26, 27]. Wang and Ding [28] used the Markov Monte Carlo Chain (MCMC) and polynomial simulation models to predict the consumption of office equipment in office buildings [28]. Author's Sun, et al. [29], Liang, et al. [30], Chung and Park [31], Mangold, et al. [32] and Williams and Gomez [33] used only statistical methods to predict energy consumption and energy efficiency in buildings: linear regression, probability density function (PDF) and similar other methods. [29-33] Zekić-Sušac, et al. [34] and Tonković, et al. [35] used four decision tree methods: classification tree (CART), conditional inference trees (CTREE), random forest (RF) and gradient boosted trees (GBT) in predicting electricity and natural gas in Croatia 's public building sector. Tonković, et al. [35] in addition to using machine

learning methods, have proposed an information system module for efficient natural gas management as a tool to assist decision makers in the Republic of Croatia in reducing natural gas consumption. [34, 35] Aside from machine learning methods, author's Mitrović and Zekić-Sušac [36] made an overview of the algorithms within machine learning methods used to reduce error where machine learning algorithms that integrated clustering algorithm were found to perform better [36]. Research papers from author's Kontokosta [37], Papadopoulos, et al. [38] and Zekić-Sušac, et al. [39] used a clustering method combined with other machine learning methods, which is also the aim of this research paper: to determine the optimal cluster procedure with the optimal architecture of the investigated machine learning methods with the least error to obtain models with high rate of accurate CO2 emissions prediction on an annual basis that should make possible to achieve greater energy efficiency in the public building sector and reduction of national CO2 emissions.

Following on from the above authors and their contributions using the methods of intelligent data analytics in the efficient management of energy and CO2 emissions in buildings, this research paper pursues to test the fundamental hypothesis and ancillary hypotheses:

1. Automated machine learning methods with minimal or no human intelligence can support decision-making through the information system to efficiently manage carbon dioxide emissions for public building managers.
    a. It is possible to develop machine learning models to predict carbon dioxide emissions in buildings that exclude or reduce human activity in the decision-making process of energy efficiency management in buildings,
    b. It is possible to incorporate Big Data technology and machine learning methods as part of an information system for efficient energy management in buildings.

The aim of this research is to analyse collected data and to propose solutions by testing hypotheses in a way by examining whether automated machine learning methods with minimal or no influence of human intelligence through the information system can support decision making for efficient management of carbon dioxide emissions for public building managers.

The scientific and social argument for further researching the topic of this paper is verified by authors Kim, et al. [40], Wang, et al. [41], Zuraimi, et al. [42] and D'Amico, et al. [43] who stated that their research had some limitations because they were making predictions in small and continuous spaces, so it „need to be validated under various building types, room arrangements, and occupants compositions to achieve a generic acceptance"[41] and that a research should have „large and reliable dataset of real and accurate designs of building structures"[43]. [40-43] In large and reliable dataset, author of this paper sees progress given that in this research there is large dataset with different types of buildings and different room layouts. Authors Hong, et al. [17] singled out as a limitation of their research failure in „presenting a regionally unified calculation model for forecasting CO2 emissions by region" [17], while the author of this paper finds it challenging to prove through methodological framework how it is possible to develop a unified calculation model for predicting CO2 emissions by region that excludes or reduces human action in the decision-making process. Authors Singh and Rao [44] propose for further development of their models to endeavour their applicability in a network and cloud environment.[44] Contributing to previous research, author of this research paper considers a necessity to integrate machine learning models for forecasting into the information system where the information system in cloud environment in its business logic would efficiently manage CO2 emissions from collected data and process to the interpretation of forecasted results. Finally the authors Kim, et al. [45] suggest that researchers in the future should consider „the four objective functions of thermal comfort, energy consumption, and economic and environmental effect at the same time"[45]. To simultaneously consider the four objective functions mentioned above is a challenging problem, therefore, the author of this paper wants to develop an improved machine learning model that would allow reliable prediction of CO2 emissions based on aforementioned objective functions.

2. Research Methods

The following scientific methods will be used to carry out the research in order to achieve the its objectives: analysis method, synthesis method, abstraction method, generalization method, analogy method, historical method, experimental method, descriptive and genetic method, statistical methods and others. The results of testing the hypotheses in previous chapter will be a synthesis of the tested fundamental and ancillary hypotheses and the development of experiment as a method.

3. Results and discussion

Taking into account results of previous research where MAPE was 2.06%, 0.160% (0.328%) in predicting CO2 emissions using evolutionary algorithms: gene expression programming - harmony search algorithm and support-vector machines [17, 18] and comparative analysis of algorithms within machine learning methods where the clustering algorithm combined with other machine learning algorithms showed favourable results - 3.443% MAPE [36], the intent is to use the building characteristics (141 attributes), their energy consumption (electricity,

gas and water) and meteorological data (temperature, pressure and wind) as input variables and CO2 emissions as output variable, to determine the optimal cluster procedure and building groups. On, by common characteristics, grouped buildings to develop models of optimal architecture of methods and algorithms, which in previous studies had low MAPE - regardless of CO2 emission or energy consumption prediction [9-31] and experiment with other methods and algorithms, such as enhanced convolutional neural networks over a ten-year time span.

4. Conclusion

Based on the results of the research, a scientific contribution is anticipated in the form of a methodological framework for predicting CO2 emissions in building sector on an annual basis integrated into an information system that by using methods and algorithms for machine learning in business logic, determines efficient management of CO2 emissions based on the characteristics of buildings, their energy consumption and meteorological data. The methodological framework itself should give the concept of CO2 emission prediction, but it should also determine the optimal architecture of the machine learning method and define the machine learning model and algorithm that will deliver the finest results. The research will be conducted on Croatia's building sector data, but its results, as well as the results of research by various authors, are generally of worldwide practicability, since from a problematic and practical point of view they should, in the near future, help achieve binding targets for CO2 reduction in the world. Attributable to the foregoing, the methodological framework for predicting CO2 emissions in building sector should be an integral part of a complete energy management information system architecture consisting of the following modules: (1) data collection, (2) data irregularity search, (3) intelligent data analytics and (4) interpretation and visualization of results.

5. References

[1] E. P. a. t. Council, "Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings," *OJ L 153, 18.6.2010, p. 13–35 (BG, ES, CS, DA, DE, ET, EL, EN, FR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV) Special edition in Croatian: Chapter 12 Volume 003 P. 124 - 146,* 2010.

[2] E. P. a. t. Council, "Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC Text with EEA relevance," *OJ L 315, 14.11.2012, p. 1–56 (BG, ES, CS, DA, DE, ET, EL, EN, FR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV) Special edition in Croatian: Chapter 12 Volume 004 P. 202 - 257,* 2012.

[3] E. P. a. t. Council, "Directive (EU) 2018/2002 of the European Parliament and of the Council of 11 December 2018 amending Directive 2012/27/EU on energy efficiency (Text with EEA relevance.)," *PE/54/2018/REV/1 OJ L 328, 21.12.2018, p. 210–230 (BG, ES, CS, DA, DE, ET, EL, EN, FR, GA, HR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV),* 2018.

[4] E. P. a. t. Council, "Decision No 406/2009/EC of the European Parliament and of the Council of 23 April 2009 on the effort of Member States to reduce their greenhouse gas emissions to meet the Community's greenhouse gas emission reduction commitments up to 2020," *OJ L 140, 5.6.2009, p. 136–148 (BG, ES, CS, DA, DE, ET, EL, EN, FR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV) Special edition in Croatian: Chapter 15 Volume 017 P. 126 - 138,* 2009.

[5] H. Tommerup, J. Rose, and S. Svendsen, "Energy-efficient houses built according to the energy performance requirements introduced in Denmark in 2006," (in English), *Energy and Buildings,* Article vol. 39, no. 10, pp. 1123-1130, 2007.

[6] K. Čulo and H. Krstić, "Cost benefit analysis of energy efficient family houses," in *Second International Conference on Harmonisation Between Architecture and Nature*, Algarve, Portugal, 2008: Eco-Architecture.

[7] N. novine, "Zakon o energetskoj učinkovitosti," *Narodne novine* vol. 2399, no. 127/2014 2014.

[8] *ISO 50001:2011 – Energy Management System,* I. O. f. Standardization, 2011.

[9] J. Wang and Q. Tao, "Machine Learning: The State of the Art," *IEEE Intelligent Systems,* vol. 23, no. 6, pp. 49-55, 2008.

[10] A. Burkov, *The Hundred-Page Machine Learning Book*, 1 ed. Kindle Direct Publishing, 2019.

[11] A. K. Jain, "Data clustering: 50 years beyond K-means," (in English), *Pattern Recognition Letters,* Article vol. 31, no. 8, pp. 651-666, 2010.

[12] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," (in English), *Science,* Review vol. 349, no. 6245, pp. 255-260, 2015.

[13] G. Shobha and S. Rangaswamy, "Chapter 8 - Machine Learning," in *Handbook of Statistics*, vol. 38, V. N. Gudivada and C. R. Rao, Eds.: Elsevier, 2018, pp. 197-228.

[14] M. R. M. Talabis, R. McPherson, I. Miyamoto, J. L. Martin, and D. Kaye, "Chapter 1 - Analytics Defined," in *Information Security Analytics*, M. R. M. Talabis, R. McPherson, I. Miyamoto, J. L. Martin, and D. Kaye, Eds. Boston: Syngress, 2015, pp. 1-12.

[15] C. McCue, "Chapter 7 - Identification, Characterization, and Modeling," in *Data Mining and Predictive Analysis (Second Edition)*, C. McCue, Ed. Boston: Butterworth-Heinemann, 2015, pp. 137-155.

[16] L. Dresner Advisory Services, "Data Science and Machine Learning Market Study " Dresner Advisory Services, LLCAugust 29, 2019 2019, Available: https://gumroad.com/l/dTfno.

[17] T. Hong, K. Jeong, and C. Koo, "An optimized gene expression programming model for forecasting the national CO2 emissions in 2030 using the metaheuristic algorithms," (in English), *Applied Energy,* Article vol. 228, pp. 808-820, Oct 2018.

[18] W. Sun and M. H. Liu, "Prediction and analysis of the three major industries and residential consumption CO2 emissions based on least squares support vector machine in China," (in English), *Journal of Cleaner Production,* Article vol. 122, pp. 144-153, May 2016.

[19] M. Kankal, A. Akpinar, M. I. Komurcu, and T. S. Ozsahin, "Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables," (in English), *Applied Energy,* Article vol. 88, no. 5, pp. 1927-1939, May 2011.

[20] J. Jeong *et al.*, "Development of a prediction model for the cost saving potentials in implementing the building energy efficiency rating certification," (in English), *Applied Energy,* Article vol. 189, pp. 257-270, 2017.

[21] C. E. Kontokosta and C. Tull, "A data-driven predictive model of city-scale energy use in buildings," (in English), *Applied Energy,* Article vol. 197, pp. 303-317, 2017.

[22] A. P. Melo, M. Fossati, R. S. Versage, M. J. Sorgato, V. A. Scalco, and R. Lamberts, "Development and analysis of a metamodel to represent the thermal behavior of naturally ventilated and artificially air-conditioned residential buildings," (in English), *Energy and Buildings,* Article vol. 112, pp. 209-221, 2016.

[23] H. Son, C. Kim, C. Kim, and Y. Kang, "Prediction of government-owned building energy consumption based on an RReliefF and support vector machine model," (in English), *Journal of Civil Engineering and Management,* Article vol. 21, no. 6, pp. 748-760, 2015.

[24] S. Farzana, M. Liu, A. Baldwin, and M. U. Hossain, "Multi-model prediction and simulation of residential building energy in urban areas of Chongqing, South West China," (in English), *Energy and Buildings,* Article vol. 81, pp. 161-169, 2014.

[25] J. S. Chou and D. K. Bui, "Modeling heating and cooling loads by artificial intelligence for energy-efficient building design," (in English), *Energy and Buildings,* Article vol. 82, pp. 437-446, 2014.

[26] S. Naji, S. Shamshirband, H. Basser, U. J. Alengaram, M. Z. Jumaat, and M. Amirmojahedi, "Soft computing methodologies for estimation of energy consumption in buildings with different envelope parameters," (in English), *Energy Efficiency,* Article vol. 9, no. 2, pp. 435-453, 2016.

[27] S. Naji *et al.*, "Application of adaptive neuro-fuzzy methodology for estimating building energy consumption," (in English), *Renewable and Sustainable Energy Reviews,* Review vol. 53, pp. 1520-1528, 2016.

[28] Z. Wang and Y. Ding, "An occupant-based energy consumption prediction model for office equipment," (in English), *Energy and Buildings,* Article vol. 109, pp. 12-22, 2015.

[29] S. Sun, K. Kensek, D. Noble, and M. Schiler, "A method of probabilistic risk assessment for energy performance and cost using building energy simulation," (in English), *Energy and Buildings,* Article vol. 110, pp. 1-12, 2016.

[30] X. Liang, T. Hong, and G. Q. Shen, "Improving the accuracy of energy baseline models for commercial buildings with occupancy data," (in English), *Applied Energy,* Article vol. 179, pp. 247-260, 2016.

[31] M. Chung and H. C. Park, "Comparison of building energy demand for hotels, hospitals, and offices in Korea," (in English), *Energy,* Article vol. 92, pp. 383-393, 2015.

[32] M. Mangold, M. Österbring, and H. Wallbaum, "Handling data uncertainties when using Swedish energy performance certificate data to describe energy usage in the building stock," (in English), *Energy and Buildings,* Article vol. 102, pp. 328-336, 2015, Art. no. 5893.

[33] K. T. Williams and J. D. Gomez, "Predicting future monthly residential energy consumption using building characteristics and climate data: A statistical learning approach," (in English), *Energy and Buildings,* Article vol. 128, pp. 1-11, 2016.

[34] M. Zekić-Sušac, A. Has, and S. Mitrović, "Recursive Partitioning in Predicting Energy Consumption of Public Buildings," in *Proceedings of the 29th Central European Conference on Information and Intelligent Systems*, Varaždin, Croatia, 2018, pp. 179-186: Faculty of Organization and Informatics, University of Zagreb.

[35] Z. Tonković, S. Mitrović, and M. Zekić-Sušac, "Business intelligence system for managing natural gas consumption of public buildings," in *36th International Scientific Conference on Economic and Social Development – "Building Resilient Society" – EBS 2018*, Zagreb, Croatia, 2018, pp. 769-778.

[36] S. Mitrović and M. Zekić-Sušac, "A systematic literature review of machine learning algorithms in modeling buildings energy efficiency," in *International Conference on Energy, Environment and Economics,*, Edinburgh, United Kingdom, 2017, vol. 5, pp. 191-196., Coventry, United Kingdom: WEENTECH Proceedings in Energy.

[37] C. E. Kontokosta, "A Market-Specific Methodology for a Commercial Building Energy Performance Index," (in English), *Journal of Real Estate Finance and Economics,* Article vol. 51, no. 2, pp. 288-316, 2015.

[38] S. Papadopoulos, B. Bonczak, and C. E. Kontokosta, "Pattern recognition in building energy performance over time using energy benchmarking data," (in English), *Applied Energy,* Article vol. 221, pp. 576-586, Jul 2018.

[39] M. Zekić-Sušac, R. Scitovski, and A. Has, "Cluster analysis and artificial neural networks in predicting energy efficiency of public buildings as a cost-saving approach," *Croatian Review of Economic, Business and Social Statistics (CREBSS),* vol. 4, no. 2, pp. 57-66, 2018.

[40] S. Kim, Y. Song, Y. Sung, and D. Seo, "Development of a Consecutive Occupancy Estimation Framework for Improving the Energy Demand Prediction Performance of Building Energy Modeling Tools," (in English), *Energies,* Article vol. 12, no. 3, p. 21, Feb 2019, Art. no. 433.

[41] W. Wang, J. Y. Chen, T. Z. Hong, and N. Zhu, "Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WiFi probe technology," (in English), *Building and Environment,* Article vol. 138, pp. 160-170, Jun 2018.

[42] M. S. Zuraimi, A. Pantazaras, K. A. Chaturvedi, J. J. Yang, K. W. Tham, and S. E. Lee, "Predicting occupancy counts using physical and statistical Co2-based modeling methodologies," (in English), *Building and Environment,* Article vol. 123, pp. 517-528, 2017.

[43] B. D'Amico *et al.*, "Machine Learning for Sustainable Structures: A Call for Data," (in English), *Structures,* Article vol. 19, pp. 1-4, Jun 2019.

[44] N. Singh and S. Rao, "Ensemble Learning for Large-Scale Workload Prediction," *IEEE Transactions on Emerging Topics in Computing,* vol. 2, no. 2, pp. 149-165, 2014.

[45] J. Kim, T. Hong, J. Jeong, C. Koo, and K. Jeong, "An optimization model for selecting the optimal green systems by considering the thermal comfort and energy consumption," (in English), *Applied Energy,* Article vol. 169, pp. 682-695, May 2016.

ISBN 978-953-6071-67-8